

**The Development of Case Assessment and Interpretation
(CAI)
in Forensic Science**

A thesis submitted to the
University of Abertay Dundee
in part fulfilment of the requirement for the degree of
Doctor of Philosophy

Graham Jackson

May 2011

I certify that this thesis is the true and accurate version of the thesis approved by the examiners.

Signed...
(Principal Supervisor)

Date. 27/05/11

List of Contents

Abstract.....	vi
Acknowledgements.....	vii
Chapter 1: Introduction.....	1
Chapter 2: Background – the state of forensic science from 1970 to 1995	
2.1 Expert opinion.....	9
2.2 Stated cases at the English and Welsh Court of Appeal	
<i>2.2.1 The case of John (Jack) Preece.....</i>	<i>10</i>
<i>2.2.2 The case of Judith Ward.....</i>	<i>13</i>
<i>2.2.3 The case of the 'Birmingham Six'.....</i>	<i>18</i>
<i>2.2.4 The case of Stefan Kiszko.....</i>	<i>20</i>
2.3 Interpretational model	
<i>2.3.1 Historical background.....</i>	<i>22</i>
<i>2.3.2 Inference and Bayes Theorem – a brief overview.....</i>	<i>23</i>
<i>2.3.3 Expressions of expert opinion.....</i>	<i>28</i>
2.4 'Value-for-money'	29
Chapter 3: A model for Case Assessment and Interpretation (CAI).....	33
Chapter 4: Refinement of the CAI model	
4.1 Introduction.....	39
4.2 The hierarchy of propositions.....	39
4.3 The difference between propositions and explanations.....	44
4.4 'Two-way' transfer.....	45
Chapter 5: The impact of Case Assessment and Interpretation on statements and reports of scientific evidence.....	48

Chapter 6: Application of the principles of Case Assessment and Interpretation to specific evidence types

6.1 Case involving DNA-analysis.....50

6.2 Illicit drugs.....54

Chapter 7: Case Assessment and Interpretation and the construction of relevant databases.....57

Chapter 8: A unifying view of the role of forensic science.....62

Chapter 9: The current status of Case Assessment and Interpretation in Criminal Justice Systems.....69

Chapter 10: Summary and Review

10.1 Introduction.....72

10.2 The case of John (Jack) Preece.....74

10.3 The case of Judith Ward.....78

10.4 The case of the ‘Birmingham Six’.....81

10.5 The case of Stefan Kiszko.....82

10.6 The case of Barry George.....84

10.7 The case of *R v T*.....87

10.8 Case Assessment.....90

Conclusions.....93

References.....97

Appendix 1: Published paper for Chapters 3

1.1 Cook R., Evett I.W., Jackson G., Jones P.J. and Lambert J.A. 1998a. A model for case assessment and interpretation. *Science and Justice*. 38: 151-156.....105

Appendix 2: Published papers for Chapter 4

2.1 Cook, R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. 1998b. A hierarchy of propositions: deciding which level to address in casework. <i>Science and Justice</i> . 38: 231-239.....	112
2.2 Evett, I.W., Jackson, G. and Lambert, J.A. 2000. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. <i>Science and Justice</i> . 40: 3-10.....	122
2.3 Cook R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. 1999. Case pre-assessment and review in a two-way transfer case. <i>Science and Justice</i> . 39: 103-111.....	131

Appendix 3: Published paper for Chapter 5

3.1 Evett, I.W., Jackson, G., Lambert, J.A. and Mc.Crossan, S. 2000a. The impact of the principles of evidence interpretation on the structure and content of statements. <i>Science and Justice</i> . 40: 233-239.....	141
---	-----

Appendix 4: Published papers for Chapter 6

4.1 Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J. and Champod, C. 2002. Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. <i>Journal of Forensic Sciences</i> . 47: 520-530.....	149
4.2 Booth, G, Johnston, F. and Jackson, G. 2002. Case assessment and interpretation in a drugs supply case. <i>Science and Justice</i> . 42: 123-125.....	161

Appendix 5: Published paper for Chapter 7

5.1 Champod, C., Evett, I.W. and Jackson, G. 2004. Establishing the most appropriate databases for addressing source level propositions. <i>Science and Justice</i> . 44: 153-164.....	165
--	-----

Appendix 6: Published papers for Chapter 8

- 6.1 Jackson, G. 2000. The scientist and the scales of justice. *Science and Justice*. 40: 81-85.....178
- 6.2 Jackson, G, Jones, S., Booth, G., Champod, C. and Evett, I.W. 2006. The nature of forensic science opinion – a possible framework to guide thinking and practice in investigations and in court proceedings. *Science and Justice*. 46: 33-44.....184
- 6.3 Jackson, G. 2009. Understanding forensic science opinions. In: J. Fraser and R. Williams, eds. *Handbook of Forensic Science*. Cullompton, Devon, UK: Willan Publishing. 2009, pp. 419-445.....197

Appendix 7: Published paper for Chapter 9

- 7.1 Jackson, G. and Jones, P.J. 2009. Case Assessment and Interpretation. In: A. Jamieson and A. Moenssens, eds. *Wiley Encyclopedia of Forensic Science*. Chichester, UK: John Wiley & Sons Ltd. 2009, pp. 483-497.....225

Abstract

Expert opinion has been, and still is, of undoubted assistance in the investigation of crime and the administration of criminal justice. It has also been the cause of a number of celebrated miscarriages of justice. Technological advances and greater investment in forensic science resources in the latter half of the 20th century meant that forensic science could be applied to many more cases than ever before. It also meant that forensic scientists had to meet new challenges in the way they formed and expressed opinions about their findings. The creation of a commercial market for forensic science in the 1990s in England and Wales put additional pressures on suppliers to provide value-for-money for their customers.

In an attempt to satisfy the potentially conflicting demands of providing robust, reliable opinion and of giving value-for-money, a novel process called Case Assessment and Interpretation (CAI), based on the underlying logic of Bayes Theorem and the use of likelihood ratios, was proposed in 1998 as a model of good practice in forensic science. Over the course of the next 12 years, the model process was applied to most main-stream forensic science disciplines and, as a result, the ideas were refined and fresh insights on the nature of expertise were gained.

This thesis describes the background to the initial development of the CAI model, sets out subsequent improvements, demonstrates how the model may have helped avoid misleading opinion being given and considers the current status of CAI. The conclusion of the thesis is that assessment of likelihood ratios, within the framework of the Case Assessment and Interpretation model, does provide a philosophical, yet practical, means of delivering robust, reliable opinion and value-for-money.

Acknowledgements

I wish to thank University of Abertay Dundee for giving me the opportunity to write this thesis and for providing resources and funding. In particular, I am very grateful to the members of my supervisory team – Professor David Bremner (Principal Supervisor), Dr Phil Collier and Dr Simona Hapca – and to Professor John Palfreyman for all their support, encouragement and advice.

I am indebted to all my colleagues in the field of forensic science, both in the UK and abroad, who have been influential over the years in shaping my thinking on Case Assessment and Interpretation. They have given their time freely, willingly and in good humour. I am very fortunate to have such valued colleagues and friends.

I would particularly like to thank members of the Case Assessment and Interpretation (CAI) team for their support, contributions and unstinting friendship over the years: Dr Ian Evett, Phil Jones, Roger Cook, Jim Lambert, Stella Jones (nee Mc Crossan), Dr Christophe Champod, Dr Gareth Booth and Dr Sue Pope. I owe a huge debt of gratitude in particular to Ian. His contributions to the underlying philosophy of forensic science have been immense. It was Ian who initially sparked my interest in, and eventual passion for, interpretation in forensic science. I value greatly his support, encouragement, advice and friendship. I have been very fortunate to have worked so closely with Ian for so many years.

My thanks are also due to Professor Colin Aitken of University of Edinburgh. It is only recently that we have worked together but I have benefitted greatly from his expert knowledge, his advice and support.

I am very grateful to Professor Paul Roberts of University of Nottingham for his advice and provision of some legal reference, and to Mr Campbell Malone, Consultant to Stephenson Solicitors, for supplying a document in relation to the case of *Stefan Kiszko*.

Finally, I am indebted to Margaret, Louise, Iain, Jamie and Grace for all their love, support and encouragement while I have been trying, over many years, to understand and help develop this thing called 'interpretation'.

Chapter 1: Introduction

The use of scientific techniques to assist criminal investigations and court proceedings is well established in many developed countries. These techniques range from relatively simple, observational techniques, such as the comparison of footwear marks, through to highly technical, analytical methods such as DNA-profiling. With the former techniques, experts tend to form their opinions from the subjective, qualitative data generated during the process of comparison of the attributes of reference and questioned samples. In contrast, the latter techniques present experts with large amounts of complex, quantitative and qualitative data generated from the analysis of the two sets of samples. Interpretation of the results of any type of technique requires expert skills that, by definition, are beyond the capability of the average lay person. While analytical techniques used in forensic science have developed extensively and rapidly over the last 30 years or so, forensic scientists and other experts have been much slower to change and improve their skills in interpretation of their findings. Furthermore, successful appeals against conviction in several high-profile criminal cases during this period were based on challenges of scientific or medical opinions that were seen as central evidence in the prosecution of the defendants. These cases posed major challenges to the forensic science community to improve not only the way in which experts formed their opinions but also the way in which opinions were communicated to lay people, police and lawyers.

Running alongside the issue of the nature and expression of expert opinion, the advent of a competitive market in forensic science in England and Wales in the 1990s forced forensic science providers (FSPs) to develop and deliver services that were more economic and more effective.

This thesis describes the body of research and development work, collectively known as 'Case Assessment and Interpretation' (CAI), that provided the forensic science community with a means of dealing with the potentially conflicting requirements of providing robust, reliable, expert opinion while, at the same time, delivering 'value-for-money' to paying customers. The mechanism proposed to deal with this conflict was a novel application to forensic science problems of Bayes' Theorem (Bayes 1763), a logical means of dealing with uncertainty through the use of subjective probabilities.

It must be stressed that the research and development work that forms the basis of this thesis, and which is described in the appended papers, was performed by a team of scientists employed by the Forensic Science Service (FSS). The initial team was composed of 'customer-facing' managers (Phil Jones and Roger Cook), forensic statisticians (Ian Evett and Jim Lambert) and a forensic science practitioner (the current author, Graham Jackson) and was a unique combination of scientists with complementary skills, knowledge and experience. The composition of the team changed over the years to meet the differing challenges that arose from the different forms of cases and issues encountered but the current author and Ian Evett were the only permanent members of this team. The author was the leader of a national project for implementation and development of 'Case Assessment and Interpretation' within the FSS.

While it is difficult to attribute each novel development to specific individuals of the team, the author of this thesis will indicate at appropriate positions in the thesis his

contribution to the body of work and, in particular, attention is drawn to the four appended papers for which the current author was the sole or lead author.

The hypothesis of the thesis is:

The evaluation of a Bayesian likelihood ratio provides a philosophical and practical model to accommodate the potentially conflicting requirements of the criminal justice system for reliable, useful, expert opinion and of customers for economic, expert services.

The Chapters of this thesis introduce and discuss, from an historical perspective, the natural phases of the development of the application of Bayes' Theorem in forensic science as described and proposed in the published papers to which this author contributed. Each Chapter deals with either a single published paper or a group of papers that form a common theme. Copies of the papers can be found in Appendices organised according to the themes of the Chapters.

The thesis is written from an historical viewpoint and, therefore, the tense used for the thesis is the past tense, even though the work described in the papers is still relevant and applicable to current forensic science practice.

The following provides an overview of the content of the Chapters.

A review of the state of forensic science from the early 1970s to the mid 1990s is provided in Chapter 2. During this critical period, the contribution and impact of forensic science within the criminal justice system increased markedly. As a

consequence, the evidence provided by forensic scientists and other experts was subjected to increasing challenge and scrutiny by the courts. In a number of high-profile cases, expert opinion was given that contributed directly to miscarriages of justice. These cases dented the previously-held impression that forensic science was highly reliable. The use of a likelihood ratio (LR), a key component within the framework provided by Bayes' Theorem, was advocated during this period by a few, key forensic statisticians as a means of providing a logical, safer approach to the interpretation of scientific evidence (e.g. Finkelstein and Fairley 1970, Lindley 1977, Evett 1983). However, assessment of likelihood ratios as a means of appraising weight of scientific evidence was not embraced enthusiastically or comprehensively by practitioners.

Also during this period, a number of governmental reviews of the arrangement of forensic science in England and Wales were commissioned, e.g. Touche Ross (1987). These reviews found that the way in which forensic science was organised and delivered was poorly connected to the requirements of law enforcement and criminal justice agencies. The 'supply' of forensic science did not meet the perceived 'demand' for forensic science services. Proposals for the re-organisation of the provision of forensic science in England and Wales were subsequently implemented in the early 1990s.

Chapter 3 focuses solely on the first paper (Cook *et al.* 1998a) in the body of work that forms the basis of this thesis. This paper introduced the notion that integration of evaluation of likelihood ratios into the structure of an operating framework would provide a robust, useful model for assessing the needs of a case as well as for

interpreting scientific findings. The resulting framework was called the 'Case Assessment and Interpretation (CAI)' model.

As a result of experience gained in applying the CAI model to real-life cases, greater understanding of the model's strengths and limitations was achieved and various refinements and developments to the model were proposed and reported in a group of three papers. Chapter 4 is based on this body of work and begins with a paper by Cook *et al.* (1998b) that introduced the novel notion of a 'hierarchy of propositions' that helped to clarify the precise contribution of the forensic science evidence in any one particular case. The important philosophical distinction between 'propositions' and 'explanations' was developed in a subsequent paper by Evett, Jackson and Lambert (2000). These two papers, combined with the first paper on the CAI model, presented an overarching structure that helped forensic scientists deliver to the criminal justice system a more relevant, more transparent and more robust service than ever before. The distinction that was drawn between propositions and explanations suggested that there were conceptually and practically different roles for forensic scientists operating either as 'evaluators' or as 'investigators' – a theme that was developed in later papers (Chapter 8). The implications of applying the CAI model in so-called 'two-way' transfer cases were explored by Cook *et al.* (1999). This paper provided, firstly, a rationale for making better decisions on what to examine within individual cases and, secondly, a better approach to evaluating the combined weight of different pieces of evidence.

Chapter 5 deals with the communication of forensic science findings through written statements and reports of evidence. The paper by Evett *et al.* (2000a) provided

guidance on how statements of evidence could be influenced by the underlying principles of CAI. The authors argued that the natural flow of the CAI model could be translated directly to the structure and content of a statement.

Chapter 6 describes the application of CAI thinking to two specific evidence types. DNA-profiling was introduced to courts of law in the late 1980s and, initially, interpretation of DNA results centred largely on estimation of a relative frequency of occurrence of the matching DNA profiles. Evett *et al.* (2002) applied the CAI model, and its probabilistic approach, to help move forward the style of interpretation. The authors also introduced the application of Bayesian networks to model the complex uncertainties involved in many DNA cases. Booth, Johnston and Jackson (2002) described an application of CAI to cases involving a specific type of drugs examination. Most drugs cases involve relatively straightforward identification and quantification of drugs but a small number of cases involve issues of 'supplying' illicit drugs. Applying the CAI model to such cases helped provide an exemplar approach to interpreting the findings in this type of case.

A vitally important area in forensic science is the collection and analysis of relevant data to inform interpretation. The paper by Champod, Evett and Jackson (2004), described in Chapter 7, provided a novel analysis of the type of data required in cases where the issue was the source of a piece of recovered material and where evidential value would be based largely on the frequency of occurrence of the material. Historically, the type of sample that was collected to provide estimates of frequencies of occurrence were usually so-called 'convenience' samples - samples that were readily and easily available. However, such samples may not be appropriate or

relevant to the issue being addressed and, as such, would not provide a robust representation of the frequency of occurrence. This paper provided guidance on the collection and construction of more appropriate data sets.

During the period from its initial development to the present date, the CAI model has been applied with varying degrees of success across all mainstream disciplines of forensic science. From that experience, fresh insights were gained about the nature of forensic science opinion, leading to a deeper understanding of the skills and knowledge required of a competent forensic practitioner. These aspects were explored in the three papers presented in Chapter 8. Jackson (2000) explored the newly-evolving concept of the role of the expert and compared that with an approach that had not changed greatly for at least the previous 40 years. Jackson *et al.* (2006) and Jackson (2009) continued to develop the notion of a distinction between the scientist operating in two subtly different roles, 'investigative' and 'evaluative', based on the philosophical basis of the type of opinion being offered. This gave fresh insights into the skills and knowledge required of forensic scientists.

Chapter 9 presents a paper by Jackson and Jones (2009) that reviewed the current status of the application of Case Assessment and Interpretation (CAI) within the Criminal Justice System and argued that forensic science practitioners cannot apply CAI thinking and actions in isolation, in a so-called 'black box' approach. For the greatest benefit to be gained, CAI has to be applied with the knowledge and engagement of all parties who commission, deliver and use forensic science within the Criminal Justice System. Without this engagement, CAI is a pointless exercise. The paper described some of the changes that are necessary within and between the

domains of forensic scientists, investigators and legal practitioners in order to achieve the perceived benefits of CAI.

Chapter 10 summarises the basics steps of Case Assessment and Interpretation and, by applying them to the stated Court of Appeal cases described in Chapter 2, demonstrate the difference that CAI may have made to these cases. The current status of CAI in today's judicial process in England and Wales is highlighted through a discussion of two recent cases in which the application of CAI was a central issue in court.

In the final Chapter, a conclusion is drawn on the central hypothesis of the thesis.

It should be noted that this thesis concentrates largely on the criminal justice system of the English and Welsh jurisdiction, this being the main domain of experience of the author. However, the principles described in the thesis can apply to any jurisdiction and, while decisions in one jurisdiction are not binding elsewhere, courts can, and do, take note of significant, relevant decisions from other countries.

Chapter 2: Background - the state of forensic science from 1970 to 1995

2.1 Expert opinion

In most of the cases in which expert opinion is provided in the form of written statements or reports, the expert witness does not attend court to provide oral evidence. There will be various reasons for this, including

- there was no suspect
- there were no proceedings against the suspect or defendant
- the defendant pleaded guilty
- the expert's evidence provided no assistance to the court
- the defence accepted the expert evidence

In a small proportion of cases, the expert does attend court and gives evidence in person. For the vast majority of these cases, there are no adverse repercussions from these appearances. Very few cases, relative to the total number that utilise forensic science, make the headlines in the reporting media and rarely does a convicted person appeal against the conviction on the grounds of misleading expert evidence. This may be taken as an indication of the robustness of expert opinion. On the other hand, an opposing view could argue that, because there is a compliant, unchallenging acceptance of the apparent strength and reliability of expert opinion, only a very small number of cases are ever challenged effectively. The recent report by the Law Commission of England and Wales (2011) highlights the ‘current *laissez-faire*’ acceptance of expert opinion evidence (page 4, 1.17).

What is undisputed is that, in a small number of high-profile cases, expert opinion was an important part of the prosecution case and that subsequent, successful appeals against conviction were based on the proposition that misleading expert opinion evidence had been given at the original trials.

The convictions of John (Jack) Preece, Judith Ward, the 'Birmingham Six', and Stefan Kiszko for crimes committed in the 1970s were dependent to a large extent on the perceived strength of scientific evidence presented at trial. The subsequent success of the appeals against conviction of each of these defendants in the 1980s and 90s was in no small part due to the appellants' lawyers challenging misleading scientific evidence, although non-scientific issues did also play a part in each of these cases. Unreliable expert opinion can be the result of poor scientific knowledge or lax application of scientific techniques. However, it can be argued that the more significant part of the problem is the way in which experts make inferences from their findings and express their opinions. Communication is a two-way process which, to be successful, relies on good 'transmission' as well as good 'reception' of the message. As such, the way in which expert opinions are understood and subsequently used, or misused, by the prosecution, defence and jury is also a significant part of the problem.

2.2 Stated cases at the English and Welsh Court of Appeal

2.2.1 The case of John (Jack) Preece¹

In 1972, Jack Preece was convicted of the murder by strangulation of a woman in the cab of his lorry in Scotland. At his trial, scientific evidence relating to the analysis

¹ *Preece v HM Advocate* Criminal Law Review 1981; 783.

and comparison of body fluids, fibres, hairs and other material was presented by forensic scientist Dr Alan Clift. No-one can say precisely what the impact his evidence had on the jury's decision but it seems reasonable to assume that it added significant weight to the prosecution's case against Mr Preece. In 1981, after questions were raised over the reliability of Dr Clift's evidence, the case was referred to the Court of Appeal. New evidence was presented at this hearing about Dr Clift's analysis of semen stains in the case and his subsequent interpretation of the results. In particular, the court heard that the victim and the defendant shared the same bloodgroup type and there was doubt that, contrary to Dr Clift's evidence, it was not possible in this case to distinguish the contribution to the analytical results of the female victim from that of the male offender (whomever he may have been). The Court of Appeal ruled:

...that had the jury heard the new evidence they must have found Dr C to be discredited as a scientific witness and that accordingly the whole of the scientific evidence he gave would have been regarded as unreliable; and the appeal must be allowed and the conviction quashed.²

There are issues in this case about the role of prosecution and defence lawyers in teasing out the strengths and weaknesses of scientific evidence. For example, it could be argued that both prosecution and defence, for different reasons, failed to help the court arrive at a fair decision on Mr Preece. Prosecution failed to ask the witness whether the bloodgroup of the victim was of the same type as the defendant. For its part, the defence failed to challenge the basis of Dr Clift's opinion that it was possible to differentiate the victim's body-fluid components from those of the offender.

² Preece (n 1) 784.

However, a key issue remained in this case – what is the duty of the expert witness in such cases and how should they discharge that duty? Commenting on the case, Brownlie refers to earlier authorities:

*The law's requirements of an expert have hitherto been that he shall give his evidence..... in a fair and unbiased manner making a full and frank disclosure so as to provide the court with the material necessary to enable it to come to a reasoned decision on the merits of the scientific issue – Davie v. Magistrates of Edinburgh, 1953 S.C. 34.*³

and, quoting Sir Roger Ormerod:

*It should be a rigorous obligation on all experts to give the court as clearly as they can the limits of accuracy of their evidence, whether it is experimental or theoretical, and to disclose, if it be the fact, that other views exist in their profession. It should also be their duty to the court to indicate what inferences cannot be drawn from their evidence.*⁴

According to Brownlie:

*For the first time it appears that the High Court is spelling out the supreme requirements for the expert, namely that he shall not only give his evidence to the best of his ability but also supply a critique of that evidence drawing attention to its weaknesses as well as its strengths.*⁵

These authorities did give useful guidance to experts on how to present opinion but they fell short of providing a unifying conceptual framework. For example, exactly

³ Preece (n 1) 784

⁴ *Scientific Evidence in Court* [1968] Crim.L.R. 240.

⁵ Preece (n 1) 785.

how should experts give evidence in an *unbiased manner*? What mechanism would direct experts to think and operate in an unbiased manner; how can experts demonstrate that they are doing that? In relation to the enjoiner '*to disclose ... other views*', it is almost certain there will be at least one other view, however unreasonable or fanciful, somewhere in the world. How can the expert know of all other views? And, finally, how does the expert decide what are the strengths and weaknesses of his evidence – how are strengths and weaknesses defined?

2.2.2 The case of Judith Ward⁶

Judith Ward was convicted in 1974 of several offences including that of causing an explosion which killed 12 people on a coach on the M62. A major plank of the prosecution case was evidence presented by a number of scientists that suggested Ms. Ward had had involvement with the handling of explosives. She herself had made certain admissions on this but her Defence contended later that those confessions were untrue. Judith Ward appealed the conviction and, at the subsequent hearing at the Court of Appeal in 1992, her conviction was quashed. In a report of the judgment, the Court comments on the scientific evidence and, in relation to the overall impact of this evidence, said:

Given that the appellant denied that she had been in contact with explosives, the scientific evidence led by the prosecution struck at the heart of the appellant's credibility. To the jury, a combination of her confessions and the supporting scientific evidence linked with all three explosions would have seemed compelling proof of appellant's guilt.⁷

⁶ *R v Ward* [1993] 1 WLR 619.

⁷ *ibid* 664.

One of the scientists who presented evidence at her trial, the late Dr Frank Skuse, had taken swabs from Ms. Ward's hands and from under her fingernails. He tested these for the presence of chemical components of explosives and he obtained positive reactions. Again, quoting from the judgment:

Dr Skuse's statement is dated February 25, 1974. It was served on the defence and used at the committal proceedings. The relevant part of the statement reads:

'A positive colour reaction for a substance similar to nitroglycerine has been obtained from each of the swabs FS3 and FS4 [fingernail scrapings] and item 47 [the ring]. No confirmatory test for nitroglycerine was obtained from these swabs. Negative tests for ammonium and nitrate were obtained from these swabs. The examination of these items is consistent with the opinion that contact of the hands with an explosive substance could have occurred. Item 45 is a brown fabric bag [the duffle bag] fitted with a string grip. The bag and grip were swabbed and positive colour reactions were obtained for a substance similar to nitroglycerine and ammonium ion a part component of ammonium nitrate. No confirmatory tests for nitroglycerine and ammonium ion were obtained. A negative test for nitrate was obtained from these swabs. The examination of these items is consistent with the opinion that contact of the inside of the bag and commercial explosive has probably occurred.'

This was not an unfair summary of the view which Dr Skuse apparently held. But it is noteworthy that at the trial Dr Skuse's evidence became more positive

*in his insistence on what he believed to be the correct inferences to be drawn from his tests.*⁸

For the sake of brevity, it is not the intention here to describe or discuss the remaining scientific evidence in this case, even though that was of considerable significance in the original trial and the subsequent Court of Appeal hearings. Consideration of Dr Skuse's evidence alone should suffice to illustrate the issues, particularly, as evidence from the same scientist occurs in the subsequent case example of the 'Birmingham Six'.

Turning to the evidence presented by Dr Skuse at the original trial, the Court says:

*Dr Skuse's evidence was along the lines of his witness statement. He insisted that his Griess test results showed that Miss Ward had probably been in contact with a commercial explosive. He also testified that a T.L.C. test of a swab taken under the left fingernail reinforced his view despite the fact that the suspect spot did not turn pink or anything like it. He was cross-examined on these points but refused to budge. If the jury chose to accept Dr Skuse's evidence, as they probably did, it followed that Miss Ward had N.G. on her person some 57 hours after the explosion at Latimer.*⁹

and later:

The prosecution did not call Dr Skuse to give evidence before us. But there is before us an impressive body of expert opinion to the effect that Dr Skuse's tests, notwithstanding his confident assertions at the trial, were of no value in

⁸ Ward (n 6) 668.

⁹ ibid 673.

establishing contact between the appellant and the explosives in 1974. The fact that Dr Skuse apparently got the Griess test result, which he described, cannot be regarded as more than an initial step towards the identification of NG: it was not evidence of the presence of NG. Dr Skuse relied on one TLC test spot, despite the fact that it was not pink. It is established to our satisfaction that this conclusion was wrong. The scientific evidence before us further shows that the interval of some 57 hours between the alleged handling of explosives at Latimer and the taking of samples by Dr Skuse rendered unlikely the suggestion of the presence of explosives on Miss Ward's hands as a result of planting explosive devices. Moreover, the very fact that the TLC test did not work puts a substantial question mark over the conditions of the preliminary Griess test. In our judgment, if the trial judge had known what we know, he would have excluded Dr Skuse's evidence as valueless. Dr Skuse's conclusion was wrong, and demonstrably wrong, judged even by the state of forensic science in 1974.

That brings us to the impact of Dr Skuse's evidence on the trial of Miss Ward. In many ways Dr Skuse was the most important scientific witness called by the prosecution. He was after all the only forensic scientist who professed to have found NG under Miss Ward's fingernails. That evidence enabled Mr Higgs to say that Miss Ward must have kneaded explosives. The elimination of Dr Skuse's evidence is therefore an important matter.¹⁰

And, finally:

¹⁰ *Ward* (n 6) 677.

*A careful study of the fresh scientific evidence has persuaded us that the scientific case against Miss Ward is now insupportable. In our judgment on this further ground Miss Ward's conviction is unsafe and unsatisfactory.*¹¹

Judith Ward's convictions on all counts were quashed.

This case provides a good example of what may have been a prevailing approach amongst forensic scientists at the time. In Dr Skuse's written evidence for the original trial, he uses the phrases such as '*... consistent with the opinion that contact of the hands with an explosive substance could have occurred*' and '*...consistent with the opinion that contact of the inside of the bag and commercial explosive has probably occurred*'. However, in his oral evidence to the court, if the reporting in the judgment is correct, he increases the weight, and therefore the impact, of his evidence by saying or implying that '*... results showed that Miss Ward had probably been in contact with a commercial explosive*'.

The scientist has progressed from what may be interpreted as a relatively mild opinion of '*consistent with an opinion that contact could have occurred...*' to a much stronger opinion of '*... had probably been in contact with....*'. This is an example of a strategy that some prosecution-commissioned scientists would use when being challenged in court: they would strengthen their opinion from that expressed in their original written report. Other scientists would adopt the opposite strategy and weaken their opinion in the face of challenge. Neither strategy is satisfactory for the effective, fair administration of justice.

¹¹ *Ward* (n 6) 679.

Whatever the motivation and skills that Dr Skuse employed as he formed the opinions that he offered in evidence, this case begs the question of whether, at this time, there was a more acceptable, more logical and robust guiding structure for forming forensic science opinions.

***2.2.3 The case of the 'Birmingham Six'*¹²**

The case of the 'Birmingham Six' also involved evidence provided by Dr Skuse and exposes similar issues to the Judith Ward case. The 'Birmingham Six', Mr McIlkenny, Mr Hill, Mr Power, Mr Callaghan, Mr Walker and Mr Hunter, were convicted in 1975 of the murder of twenty-one people in bombs that exploded in two Birmingham city centre pubs in 1974. The prosecution case relied primarily upon two strands: the confessions of the defendants and the scientific evidence apparently implicating Mr Hill and Mr Power. A key part of the scientific evidence was provided by Dr Skuse. It consisted of positive reactions for nitroglycerine obtained from hand swabs taken from Mr Hill and Mr Power. Verbatim transcripts of his evidence at the original trial are difficult to obtain and there is a distinct lack of authoritative references to this case in the publically available literature. Therefore, a secondary source (Mullin C. 1990) has been used for quotes from Dr Skuse's oral evidence. Mullin (1990) reports the following:

He (Dr Skuse) told the court that, on the basis of the Griess tests alone, he was 'quite happy' that Power and Hill had been in contact with commercial explosives.

'What do you mean by "quite happy"', he was asked.

¹² *R v McIlkenny* [1991] 93 Cr.App.R. 287.

'Ninety-nine percent certain' was his reply.

The judge at the original trial, Mr Justice Bridge, is reported as having described the forensic science evidence as one of two 'absolutely critical' chapters in the evidence against the six men. The jury, in turn, found the evidence compelling and convicted all six men. The men appealed unsuccessfully in 1976 against their convictions. Their second appeal against conviction in 1987 again failed, despite new evidence being adduced by the defence that suggested the scientific evidence was, at best, questionable or, at worst, unreliable and misleading. Finally, their third appeal in 1991 was successful - the evidence of Dr Skuse was successfully challenged following similar arguments to those in the Judith Ward case.

Of particular note in this case were the comments made by the judge at the original trial on the way in which the jury could resolve differences of opinion between experts. As quoted by the Court of Appeal at the 1991 appeal:

*Members of the jury, the resolution of scientific argument of this sort is difficult, particularly difficult for a jury of lay people, and I say once again that I am not going to try and go into the technicalities in detail because I would be in grave danger of misleading you. The only way that you can resolve these differences is by your impression of the witnesses. Use any technical knowledge that you have, but I suspect that in the end you will judge it primarily by your impression of the witnesses, and secondly perhaps by a comparison of their relative experience.*¹³

¹³ *McIlkenny* (n 12) 8.

The advice to the jury to rely on 'impressions' of the experts is perhaps understandable, given the way in which the experts' evidence was adduced and challenged in court, but it is a very poor substitute for a rational appraisal of the basis and robustness of the experts' opinions. Quite clearly, the judge was in difficulty in advising the jury arguably because he had no unifying framework with which to test the opinions.

The case raises again the question of how the experts could have presented their findings in a more reliable, more robust, more accessible way to help the jury reach its decision.

2.2.4 The case of Stefan Kiszko¹⁴

The late Stefan Kiszko served 16 years in prison after he was convicted in 1976 of the sexual assault and murder in 1975 of Lesley Susan Molseed, an eleven-year old schoolgirl, whose body was found on Rishworth Moor in West Yorkshire. A forensic scientist, Mr Ron Outteridge, gave evidence about semen stains found on Lesley's clothing. These semen stains, as would be expected from the semen of a fertile man, contained spermatozoa. However, there was a suggestion that a 'sample', presumably of semen, obtained by the police from Mr Kiszko did not contain spermatozoa. Indeed, there was medical information that Mr Kiszko suffered from a condition known as hypogonadism that affected his ability to produce spermatozoa. Whether or not Mr Outteridge knew this at the time he gave evidence is a moot point and one that is difficult to ascertain without the original trial transcripts.

¹⁴ *R v Stefan Kiszko*, CA Transcript 2665/W3/91, 18 February 1992

Mr Kiszko's first appeal in 1978 against conviction failed; his second appeal was heard in 1992. After hearing new medical evidence, Lord Chief Justice Lane said:

*It has been shown that this man cannot produce sperm. This man cannot have been the person responsible for ejaculating over the girl's knickers and skirt, and consequently cannot have been the murderer.*¹⁵

Mr Kiszko was accordingly exonerated of the crime.

A DNA profile was eventually obtained from semen from Lesley's clothing and, when entered onto the national DNA database, revealed a match with Ronald Castree. Mr Castree was subsequently charged with Lesley's murder and was found guilty of the crime in 2007.

The issue in this case is one of profoundly poor interpretation of the value of the scientific findings. If Mr Outteridge had had a robust interpretational model that would accommodate the conditioning background information of Mr Kiszko's fertility, then perhaps he would not have presented evidence that was used to convince the jury that an innocent man had committed a serious crime. In the absence of the trial transcript, it is impossible to know for certain what Mr Outteridge said in court but, judging by common practice of the day, it is likely he simply presented his finding, that of the presence of semen on the clothing, and left interpretation of its value for the lawyers to argue and for the jury to decide.

¹⁵ *Kiszko* (n 14) 10.

These four cases provide examples of poor technical knowledge and poor scientific method. What they also provide is examples of poor reasoning – reasoning that leads to opinions that are not robust, not transparent and not balanced and, therefore, highly prejudicial against defendants who eventually were found to be innocent in the eyes of the law.

2.3 Interpretational model

2.3.1 Historical background

Until the 1980s, scientists had very little formal training in the skills of interpretation beyond that of 'learning by apprenticeship' – trainee scientists would work alongside more experienced colleagues, watch what they did and listen to the opinions expressed. These expressions of opinion would comprise a wide range of words and phrases to convey the scientists' views on the meaning or strength of their findings (see later, section 2.3.3). Furthermore, there was nothing in forensic science training in the UK generically on inference or on dealing logically with uncertainty. Statistical training was limited and consisted mainly of the application of significance tests in, for example, the comparison of measurements of physical and chemical properties of glass fragments. Evett (1991) provides a very good summary of this situation in the 70s and 80s, and Aitken (2009) describes the progression of the use of statistics in forensic science from 'relative frequencies' through 'discriminating power' and 'significance probabilities' to 'likelihood ratios'. There had been early papers exploring the application of Bayes' Theorem to forensic science problems, including one by Finkelstein and Fairley (1970) who were based in the United States of America. However, it was a paper by a UK-based statistician (Lindley 1977) that stimulated Evett, a forensic scientist and statistician working for the Forensic Science

Service in England and Wales, to push forward with applying Bayesian thinking to the interpretation of comparison of measurements on questioned and reference samples of not only glass but of all transfer evidence. To quote Evett (1991, page 12):

What is the functional relationship between coincidence probability and evidential strength? Dennis Lindley's response (1977) was to show how the Bayesian likelihood ratio ... provides a conceptually simple and coherent one-stage solution to the problem.

and later (page 12):

Thus evidential value is measured by a continuous function with no arbitrary cut-off.

For the first time, with this reference, there is the germ of guidance on how scientists should appraise logically the strength of their findings.

2.3.2 Inference and Bayes' Theorem – a brief overview

Triers of fact in Courts of Law are required to consider the evidence that is presented to them and to use that evidence in reaching a decision on an uncertain event – the guilt or innocence of the defendant. One would hope that the triers of fact would reason carefully and apply, knowingly or not, the rules of logic to reach their inferences. Much has been written on inference and logic, but most authorities recognise three forms of inference: deduction, induction and abduction. Deduction is necessarily a deterministic process – the inference flows naturally from universally true premises. It is sometimes described as arguing from the general to the specific. Examples of deductive inferences in forensic science would include the answers

provided by classification systems and the exclusion of samples where questioned samples do not match the putative source. Inductive inference is not deterministic; it is ampliative, in the sense that the inference adds to our knowledge but does not provide a conclusive answer. The truth of the major premise of the argument will be advanced or pushed back as each new piece of information or evidence is considered. Inductive inference requires an understanding of the numerical associations between observations and therefore relies on statistics about these observations. Inductive inference is also described as arguing from the specific to the general and is entirely relevant to the issues of importance within the judicial process. Finally, abduction is the generation of hypotheses to explain observations; it has a key role in investigative situations where there is uncertainty about the circumstances of an incident. Anderson *et al.* (2005, page 55) provide a succinct summary of these three forms of inference.

Deduction, induction and abduction all have their part to play during the course of the investigative and judicial processes but, when considering at trial the evidential force or weight provided by the scientific observations (evidence), induction is the form of inference that is best suited to providing that assessment. Furthermore, many scholars of legal reasoning advocate a probabilistic approach to inductive logic. Feinberg (1997), commenting on the various forms of reasoning that have been proposed, asserts:

“... to remedy the claimed inadequacies of the theory of probability, in recent years we have been exposed to such alternatives as Cohen’s Baconian inference, Dempster-Shafer belief functions, fuzzy sets etc. In the end, when

subjected to careful scrutiny, each of these alternatives has been tried and found severely wanting.”

and later:

“This does not mean that probability theory, and with it Bayes’ Theorem, is a panacea for all situations.”

The current author agrees fully with these comments and can add that he has not seen any method, other than evaluating a likelihood ratio within a Bayesian framework, for assessing the weight of scientific evidence in a coherent, robust, accessible way for the court.

The Reverend Thomas Bayes, an 18th century cleric and mathematician, is generally attributed with first describing the Theorem - a mathematical formula that demonstrates how uncertainty about a possible event can be updated in a logical way when new information or evidence is presented (Bayes 1763). The probability of event (H) given the background information (I) is denoted by $\Pr(H|I)$ and represents the prior uncertainty, or *prior probability*, of that event. The vertical bar ‘|’ in the notation is called the conditioning bar and can be read as ‘given’ or ‘assuming’. The probability in question is dependent on, is informed by, or conditioned on, the background information. The probability of event (H) given the new evidence (E) is denoted by $\Pr(H|E,I)$ and represents the updated uncertainty, or *posterior probability*, of that event. Bayes’ Theorem demonstrates the *posterior probability* is proportional to the product of the *prior probability* and the *probability of the evidence* given the truth of the event:

$$\Pr[H | E, I] \propto \Pr[E | H, I] \times \Pr[H | I] \quad \text{[Equation 1]}$$

The middle term in this expression, $\Pr[E|H,I]$, is known as a *likelihood* and, although superficially similar to the *posterior probability*, it is crucially different – it is a reversal of the terms of the posterior probability. Instead of the probability of the event in question, given knowledge of the new evidence, the *likelihood* is the probability of obtaining the new evidence given that the uncertain event were actually true.

In situations where two mutually exclusive, exhaustive events are being considered, then the so-called Odds Form of Bayes' Theorem can be applied. In this form, the *posterior odds* of the event are equal to the *likelihood ratio* for the evidence multiplied by the *prior odds* of the event. In judicial situations where there is a prosecution proposition and a defence alternative, then the Odds Form of Bayes' Theorem can be represented as follows:

$$\frac{\Pr(H_P | E, I)}{\Pr(H_D | E, I)} = \frac{\Pr(E | H_P, I)}{\Pr(E | H_D, I)} \times \frac{\Pr(H_P | I)}{\Pr(H_D | I)} \quad [\text{Equation 2}]$$

where H_P and H_D represent respectively the prosecution and defence positions.

The right-hand ratio of this equation represents the *odds* of the prosecution proposition being true, prior to the scientific evidence being heard or adduced. This expression is therefore known as the *prior odds*. The left-hand ratio of the equation, following similar nomenclature, is known as the *posterior odds* of the prosecution proposition being true, after the scientific evidence has been presented. The *posterior odds* are then the product of the *prior odds* and the ratio in the centre of the equation – that ratio is called the *likelihood ratio*. In words, it can be expressed as: the probability of obtaining the evidence, given that the prosecution proposition were true, divided by the probability of obtaining the evidence, given that the defence

proposition were true. The *likelihood ratio* can be viewed as a measure of the value of the evidence. In other words, it is a measure of the impact of the scientific evidence on the *prior odds*. It is the *likelihood ratio* that provides an elegant means of interpretation of evidence. An appraisal of the value of the likelihood ratio requires the scientist to draw upon all the expert knowledge that they possess and it provides the framework within which the scientist operates. It is truly the domain of the scientist.

It is important to note that the background information 'I' will consist of those, and only those, parts of the case circumstances that influence, or condition, the probability assignments. However, the parts of the background information that condition $\Pr(H_P|I)$ and $\Pr(H_D|I)$ will be different from those parts that condition $\Pr(E|H_P, I)$ and $\Pr(E|H_D, I)$. This is a very important notion when considering so-called 'context effects' (Saks *et al.* 2003). These will be introduced and discussed later in Chapter 3 of this thesis.

Despite publishing numerous seminal papers (Evetts 1983, Evetts 1984, Evetts 1986, Evetts 1987a, Evetts 1987b, Evetts 1989, Evetts 1990, Evetts 1991) covering both theory and practical examples, Evetts's proposal of Bayesian inference, and, in particular, appraisal of a likelihood ratio, as a unifying conceptual framework for forming and utilising expert opinion was not embraced by the profession in the United Kingdom. As late as 1996, a paper by Rudram (1996) and ensuing correspondence (Taroni and Aitken 1996, Davis *et al.* 1996, Aitken and Taroni 1996) revealed that there was still misunderstanding and controversy over the so-called 'Bayesian approach' among leading professional forensic scientists. There will have been various reasons for

practitioners' reluctance to adopt a likelihood ratio approach and, in the view of the current author, these will have included:

- aversion to changing methods that appeared to have worked well over the course of many years;
- difficulty in understanding Bayes' Theorem;
- dislike of using likelihood ratio assessments instead of offering posterior probabilities for propositions or, more simply, offering explanations.

2.3.3 Expressions of expert opinion

It can be suggested that the way in which individuals express their opinions provides a window on their thought-processes. Words can reveal how scientific findings have been appraised and how inferences have been formed. Both Rudram's work (1996) and earlier studies by Craddock, Lamb and Moffat (1989) and Satterthwaite and Lambert (1989) found a wide range of phrases being used by scientists to express their opinions. For some of these phrases, there was a wide variation between scientists and their clients (police, lawyers etc) in the perception of the strength of the phrase. Furthermore, even the within the scientists' group, there was wide a variation in scientists' perception of the strength of some phrases. Although the authors did not attribute the 'within-scientist' variation to any particular factor, it could be argued that apprenticeship-style of teaching, as mentioned previously, would influence trainee scientists to emulate their individual mentors, who would have their own particular way of thinking and their own habitual, favourite way of writing opinions.

A further feature to note about experts' statements of the time, based on the current author's personal knowledge, was the brevity of such statements, with little

explanation of the rationale of the examinations or of how opinions had been formulated.

To summarise, during this critical period there was significant scope for improvement firstly in the way experts formed and expressed opinions, secondly, in the consistency of opinions between scientists and, finally, in the mutual understanding between providers and users of expert reports and statements.

2.4 'Value-for-money'

Government initiatives in the late 1980s encouraged public services to be more cost-effective and efficient in their operations and service-provision. In some areas, the means of achieving such aims was through the introduction of direct charging of customers for the service provided by the public body. Forensic science was one such area that, up to that date, had been centrally funded from local and national government, with very little connection between the supply of services and the demand for those services. A key report by Touche Ross (1987) contained proposals for the re-organisation of the provision of forensic science along more business-like lines. This report kick-started the process of transformation of the provision of forensic science in the English and Welsh jurisdiction and, with the subsequent introduction of payment directly by customers for each case examination or each service, attention of both suppliers and customers turned to achieving 'value-for-money'. The basic notion underpinning the transformation to a 'business-like' organisation was that, by giving customers the freedom and the means to buy whatever forensic services they wanted, customers would focus on identifying and purchasing only the most cost-effective services on offer. This, in turn, would drive

providers to improve the quality and quantity of the services offered. Gallop (2003), Fraser (2007), Tilley and Townsley (2009) and Cooper and Mason (2009) all provide succinct historical accounts of the position of forensic science in England and Wales at that time and the changes that ensued in the relationship between the police as 'customers' and the forensic science organisations as 'providers'. Fraser and Williams (2009) give an overview and comparison of the political, social, economic and legal environment of forensic science in the 1980s to the present date. Jackson and Jones (2009) provide an historical account of the background to the development of a model for Case Assessment and Interpretation (CAI). They describe how assessment of likelihood ratios for the potential outcomes of any proposed scientific examinations would provide the philosophical and practical basis to meet the requirement of providing 'value-for-money'.

One of the key changes in the relationship between suppliers and customers in the early 1990s was the introduction of Scientific Support Managers in police forces in England and Wales. Part of this new role was to manage the police budget for forensic science. Perhaps inevitably, systems for monitoring the costs and the benefits of forensic science were introduced. While costs in the form of monthly spend were easy to measure, benefits were much harder to ascertain. Performance measures such as speed of delivery, number of items examined and number of 'positive' cases were relatively straightforward to implement but, by themselves, did not get to the heart of measuring the effectiveness of forensic science. A major hurdle was the problem that the output of forensic science could have an impact at any point along the investigative and criminal justice processes from incident through to court. It was hard, therefore, to tease out from other, non-forensic inputs exactly what

contribution was made by forensic science. Fraser (2000) gives a flavour for the complexity faced by police forces when trying to evaluate the effectiveness of the services of forensic science providers. Tilley and Townsley (2009) provide a review of various studies and reports on the effectiveness of forensic science. They refer to earlier work by Tilley and Ford (1996) that evaluated the use by police forces of forensic science. One of the findings of this earlier work was that (page vi):

Current within-force routine methods of estimating the effectiveness of forensic related work have dubious reliability or validity. They are, at best, starting points for further investigation.

McCulloch (1996) provides a comprehensive description of the data on 'forensic support' collected by a sample of police forces and makes a number of recommendations for developing a more refined system, including a suggestion that “*a central body should be tasked with producing ... more definitive definitions for evaluation scores*”. To date, the current author is unaware of such a development.

Forensic science providers responded in various ways to facilitate the new relationship with customers and to drive the internal changes necessary to make that relationship work. For example, the main provider of forensic science in England and Wales, the Forensic Science Service (FSS), introduced managerial roles that were ‘customer-facing’ and implemented a system for monitoring costs and effectiveness. However, despite entering into joint projects with the customers, the FSS could only assess effectiveness from their limited standpoint as a supplier.

It could be argued that the late 1980s and early 90s saw a conjunction of various pressures on forensic science. Firstly, improvements in technology, and increasing reliance on forensic science for investigations and criminal trials, meant there was ever-increasing demand on forensic science providers. Potentially huge resources would be required to examine all the items that could possibly be collected from scenes of crime and from suspects; however, there are only ever finite, limited resources available to carry out those examinations. Decisions had to be made on which items to examine and which tests to employ in any one particular case. Secondly, expert opinion was widely seen to be biased in nature and over-emphasised in value. It appeared that only a paradigm shift in the underlying philosophy of forensic science would resolve these pressures. The shift would not be simply esoteric; it needed to be a shift in forensic science thinking and practice that would impact on the whole investigative and criminal justice process.

In response to these pressures, ‘customer-facing’ managers in the Forensic Science Service (FSS), Phil Jones and Roger Cook, engaged with FSS forensic statisticians, Ian Evett and Jim Lambert and an FSS forensic science practitioner, the current author, Graham Jackson, to form a unique team, the CAI Team, to develop a rational, effective framework to guide practitioners. The project required the complementary skills and knowledge of the team members, but also relied heavily on the outputs of facilitated workshops, engaging with the majority of FSS’ practitioners, on case studies across all scientific disciplines in the FSS. The current author was the facilitator at these workshops and engaged with practitioners across all disciplines to test the initial framework, to gain new insights and to refine the framework. These refinements are described in Chapters 4 to 8 of this thesis.

Chapter 3: A model for Case Assessment and Interpretation (CAI)

The changing environment of the Criminal Justice System in England and Wales in the 1990s stimulated the development of a new relationship between the providers of forensic science and the police, prosecution and defence authorities, who were now paying customers. The pressing requirement to provide value-for-money in forensic science, while still ensuring that the service was reliable and robust, offered potentially conflicting demands that needed to be accommodated. To that end, the FSS' development team, described at the end of the preceding Chapter, set itself the task of developing a model that would:

“...enable decisions to be made which will deliver a value for money service meeting the needs of our direct customers and the Criminal Justice System.”

The model that was subsequently developed and described by Cook *et al.* (1998a, provided in Appendix 1.1) was called the Case Assessment and Interpretation (CAI) model and addressed the issue of how to make informed, rational, effective decisions on which items to examine and which tests to use in individual cases. The authors based their proposal on the logical framework provided by Bayes' Theorem, with a focus on assessment of the likelihood ratio to indicate potential evidential value and to help ensure any opinion formed from the findings was balanced, robust and justified. Lawless (2010) described the CAI model approach as ‘*a notable attempt to reform the role and position of forensic scientists within the current criminal justice system.*’

The model for CAI was depicted as a staged process composed of three phases of work:

- customer requirement
- case pre-assessment
- service delivery

While the third stage - 'service delivery' - was broadly what all forensic scientists would normally do in some form or other, the novel and critically important feature of the model was the description of two additional phases – 'customer requirement' and 'case pre-assessment' – that should be carried out by the scientist before they embarked on delivering their usual service. Conventionally, forensic scientists would be presented with a set of items to examine along with some form of request or instruction from the police. At one end of the spectrum, this request could be phrased in a broad form, such as '*... provide forensic evidence against the suspect*'. With this form of request, there is scope for the scientist to decide which items (exhibits or productions) to examine and which tests to use. At the other end of the spectrum, the instruction could be in the form of a specific direction. For example, the request could be along the lines of '*... examine the clothing from the accused for glass fragments and compare with the control sample from the broken window.*' Even with this example, while there may be limited scope for decisions to be made on which items to examine, there may still be decisions to take on which tests to employ when comparing any glass fragments that may be recovered. The crucial question in both situations is one of how were these decisions to be made and, when they were made, were they good decisions?

Research by Ramsay (1987) and by Tilley and Ford (1996) showed that police officers, generally, were ill-informed about the use and benefits of forensic science. A proportion of police requests for forensic science examinations were based on poor

understanding of the potential benefits and limitations of the service being requested. On the other hand, scientists had very limited understanding of the circumstances of the case being submitted and, through tradition or through a perceived pressure to complete the examination, tended not to try to elicit any more information from the police. Furthermore, scientists employed only those tests and techniques that were available in their own laboratory. The deployment of tests and techniques was based on a degree of logic about, for example, the specificity and sensitivity of the tests, but was also based on tradition and heuristics.

The first phase in the process of the CAI model, the so-called ‘customer requirement’ phase, encourages the scientist, in each case, to engage pro-actively with the police (or, more widely, the customer) and, through dialogue, gain an understanding of the detailed needs of the case. The scientist should aim to understand the issues that were in dispute, or the issues for which there was uncertainty. It would be only through this process that the scientist could then start to develop an examination strategy of which items to examine and which tests to use to help address the issues.

In the second phase of the model, the phase of ‘case pre-assessment’, attention is focussed on assessing the most likely outcomes of the various examination options, along with the potential evidential strength of each outcome. The means of achieving this assessment is through, firstly, definition of appropriate pairs of propositions that reflect the prosecution and defence positions. Secondly, following Bayes’ Theorem, the scientist should consider and assign probabilities or, more precisely, likelihoods for the potential outcomes of the examinations, given the truth of both the prosecution proposition and the defence alternative, and given the conditioning background

information 'I'. The only factors within the background information that should influence the scientists' assignments would be those aspects of the circumstances that have a bearing on the likelihoods of obtaining the different outcomes. These factors would include, for example, what witnesses say about the activity carried out by the offender and what the defendant offers as an alternative. Saks *et al.* (2003) suggest there is 'domain specific' and 'domain irrelevant' information and give examples of differing situations in which 'domain irrelevant' information affects adversely the interpretation of the scientist. The authors call this adverse influence the 'context effect'. In the CAI model, when assigning likelihoods at the pre-assessment stage, the scientist is 'blind' to the eventual outcomes of their examinations and is therefore uninfluenced by those outcomes. In addition, by focussing on those, and only those, parts of the background information that condition likelihoods, it seems that only 'domain specific', not 'domain irrelevant', information is being considered. Indeed, Saks *et al.* (2003) mention (page 88) the pre-assessment phase of CAI as a suitable means of reducing the risk of 'context effect'.

As described in Section 2.3.2 of this thesis, the ratio of the two likelihoods is known as the likelihood ratio (LR), the broad magnitude of which provides an assessment of the weight of evidence for each specific outcome of the examination. Not only that, the pre-assessment also provides the probability of achieving these outcomes if the suspect is truly guilty or truly innocent, thereby providing the customer with an assessment of likely usefulness of the examination.

On completion of this second phase of the model, an analysis of the customer-requirement and of a proposed examination strategy could be presented to, and

discussed with, the customer who, as the commissioner of the work, has the final say on whether or not to proceed with the examination.

Establishing the customer-requirement and devising effective examination strategies require very good communication and engagement between supplier and customer. If done well, this process would provide the best chance of delivering an effective and economic service and, thereby, help to reduce the amount of wasteful, unproductive work that may have been commissioned. Effective application of the first two phases of the model would help save costs to both customer and provider, and ensure that forensic science addressed the agreed key issues in each case, not just those issues that appeared to the scientist to be important or to the customer as the most obvious in the circumstances. The logical basis of Bayes' Theorem provides the essential framework to guide practice.

It was recognised by the authors that the most effective forensic scientists had routinely practised, arguably unknowingly, the first two phases of the CAI model but that process had never been explicitly formulated or stated prior to this paper. For the first time, 'best practice' had been analysed, described and made transparent. As a result, forensic scientists, police and lawyers now had a published model that could be challenged, tested and improved. Furthermore, for the first time in forensic science, the two potentially conflicting pressures – the requirement from the criminal justice system for robust, reliable, impartial, expert opinion and the drive from paying customers for effective, efficient, economic services from forensic science providers – had been accommodated. The contribution of this paper to the practice of forensic science was recognised by an award for 'Referenced Best Article' from the European

Academy of Forensic Sciences in 2000 (ENFSI 2000) and, more recently, the model formed the basis of standards for forming evaluative opinion written by a consortium of all the main forensic science providers in the United Kingdom and the Republic of Ireland (AFSP 2009).

Chapter 4: Refinement of the CAI model

4.1 Introduction

The CAI model is founded on the logical framework of Bayes' Theorem and, as such, there can be little argument against the central tenet of the model. However, when the CAI model was first described, there had been limited application of Bayesian thinking, and therefore of the model, to real-life forensic science cases. The authors were well aware that wider exposure and application of the model across the broad range of forensic science disciplines would reveal the benefits and limitations of the model. Over the course of the next four years or so following the initial publication, the authors embarked on a series of workshops with their practitioner colleagues and, to a more limited extent, with a small selection of police, lawyers and judges, to explain the model, to apply the model to a wide variety of case types and to gain feedback to help develop the model. As a result of this work, a number of papers were published that described the developments and the new insights that were gained. This chapter focuses on a set of three of those papers (see Appendix 2) that deal with developments that were of a more generic nature.

4.2 The hierarchy of propositions

The paper by Cook *et al.* (1998b), provided in Appendix 2.1, introduced the novel notion of a 'hierarchy of propositions' that helped to formalise and clarify the precise contribution of the forensic science evidence in any one particular case.

Perhaps the single most important element of the CAI model is the definition of a pair of propositions from which likelihood ratios for scientific findings can be derived. In any one particular case, propositions could be phrased by scientists in a variety of

ways, all of which seemed relevant to the case. Analysis of these different forms of proposition revealed there was a hierarchical structure that connected the propositions in a logical way and that seemed to be generic in application. At the ‘lowest’ level in the structure, the scientist would consider propositions about the source of the ‘questioned’ material that they had recovered. Examples would include propositions such as

- *the fibres recovered from the surface of the car seat came from the suspect’s pullover;*
- *the blood on the shoe of the defendant came from the victim of the kicking assault;*
- *the footwear mark at the scene of the crime came from the shoe of the accused.*

The value of the findings in these situations would be related usually to an assessment of the ‘within-sample’ and ‘between-sample’ variability to inform probabilities for, respectively, the numerator and denominator of a likelihood ratio. This level of the hierarchy was designated level I and is called *source-level*.

At the next level in the hierarchy, designated level II, the scientist would be considering propositions relating to an activity that the accused is suspected of doing.

Examples would include:

- *the suspect drove the stolen car at the time of the robbery;*
- *the defendant was the person who kicked the victim;*
- *the accused trod on the bathroom floor.*

Assessment at this so-called *activity-level* automatically takes into account considerations at source-level and incorporates additional factors such as the amount and distribution of the apparent transferred material. The scientist should be best-

placed to offer expert knowledge of transfer and distribution and, therefore, best-placed to offer assistance to the triers of fact in relation to issues of whether or not the accused actually did the activity.

The final level in the hierarchy, level III, involves consideration of propositions relating to the crime that had been committed and was duly termed *offence-level*.

Examples of this level would include:

- *the suspect committed the robbery;*
- *the defendant assaulted the victim;*
- *the suspect was the burglar*

Analysis at offence-level would necessarily include consideration of relevant activity-level propositions. However, the analysis requires additional consideration of issues such as intent or, as in the case of some sexual offences, consent. The scientist is generally prevented from addressing offence-level propositions because issues of intent and consent are entirely in the province of the triers of fact and are completely outside the knowledge, expertise and role of the scientist.

By focussing on activity-level propositions, the scientist is encouraged to provide more effective opinion than simply staying at source-level. Activity-level is closer to the ultimate issue of guilt/not guilt and is a domain in which the scientist can bring to bear all their knowledge and expertise about transfer, persistence, distribution, detection and background levels of the evidential material in order to help the court arrive at a decision.

It should be noted that, with the three examples of activity-level propositions given earlier, there has to be an assessment of the probability that the suspect/accused was wearing the items at the time of the crime. In some situations, it may be that the scientist is able to make an explicit assumption that the suspect was wearing the item at the relevant time and, of course, this would be open to challenge. In other situations, there may be more uncertainty about the issue of wearing and this may be best left for the court to appraise. If that were the case, then the scientist could consider modified propositions such as:

- *this is the clothing worn by the person who drove the stolen car at the time of the robbery;*
- *this is the clothing worn by the person who kicked the victim;*
- *this is the shoe that was worn by the person who trod on the bathroom floor.*

Finally, consideration of activity-level propositions enables the proper appraisal of the value of the absence of evidence. If no transferred material has been found, and the scientist operates simply at source-level, then all that the scientist can say essentially is that they have found nothing. If the scientist operates at activity-level, then the absence of transferred material could well provide support for the defence alternative – something that cannot be achieved if the scientist stays at source-level.

The current author's key contribution to this paper was the recognition and naming of the various levels of the hierarchy referred to as the 'hierarchy of propositions'. For the first time, practitioners had a framework within which to set their work and thereby to help define the purpose of their work.

As an indication of the impact of this paper, it has been cited in a major report by the National Academy of Sciences on the provision of forensic science in the United States (National Academy of Sciences 2009, p117-118). The following is reproduced from the report, with the original footnote included:

“Analyses in the forensic science disciplines are conducted to provide information for a variety of purposes in the criminal justice process. However, most of these analyses aim to address two broad types of questions: (1) can a particular piece of evidence be associated with a particular class of sources? and (2) can a particular piece of evidence be associated with one particular source?”

and later ...

“Although the questions addressed by forensic analyses are not always binary (yes/no), the paradigm of yes/no conclusions is useful for describing and quantifying the accuracy with which forensic science disciplines can provide answers.⁵”

“⁵More complete discussion of the questions addressed by forensic science may be found in references such as K. Inman and N. Rudin. 2002. The origin of evidence. Forensic Science International 126:11-16; and R. Cook, I.W. Evett, G. Jackson, P.J. Jones, and J.A. Lambert. 1998. A hierarchy of propositions: Deciding which level to address in casework. Science and Justice 38:231-239.”

It is significant that this paper is one of the few European papers quoted in the NAS report, illustrating that the ideas developed in the paper have been acknowledged in the United States, a country that, to date, has been reluctant to adopt any form of Bayesian thinking in forensic science.

4.3 The difference between propositions and explanations

The important philosophical distinction between 'propositions' and 'explanations' was explored and developed in the paper by Evett, Jackson and Lambert (2000), provided in Appendix 2.2. This paper laid the groundwork for an evolving idea that there were potentially two different roles for forensic science depending on whether the scientist was considering pairs of propositions, before they performed any substantive examination, or was offering alternative explanations after they had obtained results of their examinations. The current author's key contribution to this paper was the recognition that, when scientists offered opinions, there was a subtle but important distinction into two classes of opinion.

When trying to define propositions, it can be relatively easy for the scientist to set out the so-called prosecution proposition along the lines of the allegations or suspicions of the police and/or prosecutors. Defining the alternative to the prosecution's proposition can also be relatively straightforward if the defendant has given an alternative story in interview or, later, in evidence. However, in those cases where the defendant has not given any intimation of their story, or has not stated their version of events, then the scientist is in a quandary. In these situations, there was a tendency for scientists to generate alternative explanations for their findings, after they knew what those findings were. This is an open-ended process with potentially a large number of alternative explanations, some of which could be reasonable while others would be fanciful or even completely unrealistic. If the scientist does generate an alternative explanation, and if this alternative were indeed true, it would follow that the likelihood of the findings would approach a probability value of 1. It may well be that the likelihood of the findings if the prosecution proposition were true would also

approach a value of 1. In that situation, the likelihood ratio for the scientific evidence would be 1 and the evidence would therefore be of no probative value. This result did not fit well with a general view that positive findings did indeed have some probative value. The key to resolving this difficulty was the requirement that the scientist should assess, before carrying out any substantive work, the potential value of the findings based on the prosecution propositions and defence alternatives as they were known at that time. This approach would avoid the charge of *post hoc* rationalisation – the construction of an alternative that fits the findings – and is another example of reducing the risk of ‘context effects’ (Saks *et al.* 2003).

There is a role, however, for the construction of such explanations and this occurs in investigative situations when, for example, there is no suspect or where there is uncertainty about what happened at a scene of a crime, irrespective of who committed that crime. Saks *et al.* (2003) give a good example (page 85) of the distinction between ‘investigation’ and what they describe as ‘forensic science’ or, what could be described more appropriately as ‘evaluation’ of weight of evidence.

4.4 ‘Two-way’ transfer

The cases used as examples in the papers cited so far involved essentially ‘one-way transfer’ of evidential material. Interpretation of the findings at activity-level in this type of case essentially requires the scientist to consider two questions of the form:

- *how likely is it that this amount of material, in this location and giving this (chemical) analysis would have been obtained if the prosecution were true?*
- *how likely is it that this amount of material, in this location and giving this (chemical) analysis would have been obtained if the alternative were true?*

However, many cases involve more than one type of evidential material and the implications of applying the CAI model in so-called 'two-way' transfer cases were explored in a paper by Cook *et al.* (1999), provided in Appendix 2.3. The authors used an example in which there was potential transfer of fibres from the clothing of a suspect to that of the victim as well as from the victim to the suspect – so-called 'two-way' transfer.

The key consideration in evaluating the weight of the evidence in such cases is that of assigning conditional probabilities for the second transfer, or second type of evidence. While the scientist will still consider the pair of questions as outlined above, they now have to consider the following additional questions:

- *how likely is it that this amount of material, in this location and giving this (chemical) analysis would have been obtained if the prosecution were true AND GIVEN the findings in relation to the first transfer?*
- *how likely is it that this amount of material, in this location and giving this (chemical) analysis would have been obtained if the alternative were true AND GIVEN the findings in relation to the first transfer?*

These additional questions expose conditional probabilities for the second set of findings and may, if assigned robustly, demonstrate limited or no additional value in performing the second examination. The potential benefit of saving time and resources would only be achieved if this type of assessment were to be done before carrying out the work rather than after the work had been completed.

This paper provided the rationale for making informed decisions on whether it was of value to examine further items or employ further tests within a case. It also offered a

more robust, logical approach to evaluating the combined weight of different pieces of evidence.

Chapter 5: The impact of Case Assessment and Interpretation on statements and reports of scientific evidence

The paper by Evett *et al.* (2000a), provided in Appendix 3.1, suggested a structure for writing statements of evidence. It was argued that the logical process, both of thought and of deed, captured by the CAI model should be reflected in the structure and content of statements written for the purposes of criminal proceedings. The CAI model was based on phases that flowed logically from each other and which led naturally to a conclusion. It was proposed that headed sections in statements and reports would capture these phases and demonstrate both the rationale for the examination and the way in which the findings had been interpreted. If done well, readers of the statement should be led to arrive at the same conclusion as the scientist even before they read that conclusion. The structure suggested by the CAI model provided transparency about what the scientist had done, why they did it and how they had arrived at their opinion.

This particular paper (Evett *et al.*, 2000a) was mentioned in the National Academy of Sciences report on forensic science in the United States (National Academy of Sciences 2006, p185-186) in a chapter entitled ‘Improving Methods, Practice, and Performance in Forensic Science’. Within this chapter, there is a section called ‘Reporting Results’ in which three papers are recommended as guidance on terminology and on interpretation of evidence. Relevant parts of the report are reproduced here, with the reference from the footnote shown in brackets:

“There is a critical need in most fields of forensic science to raise the standards for reporting and testifying about the results of investigations.”

and later ...

“Although some disciplines have developed vocabulary and scales to be used in reporting results, they have not become standard practice. This imprecision in vocabulary stems in part from the paucity of research in forensic science and the corresponding limitations in interpreting the results of forensic analyses. Publications such as Evett et al.⁵, Aitken and Taroni⁶, and Evett⁷ provide the essential building blocks for the proper assessment and communication of forensic findings.”

(⁵ I.W. Evett, G. Jackson, J.A. Lambert, and S. McCrossan. 2000. The impact of the principles of evidence interpretation on the structure and content of statements. Science and Justice 40(4):233-239.)

All the major providers of forensic science services in the United Kingdom follow the basic structure proposed in the paper. The current author is also aware that providers in the Republic of Ireland, Holland and Sweden are considering, or are trying out, a similar structure.

A further, novel development described in this paper was the introduction of a fourth level in the hierarchy of propositions. The current author was conscious that there was a problem with the evaluation of evidence in those cases in which DNA-profiles could not be attributed to specific body fluids or tissues. He introduced the novel notion of sub-level I, or *sub-source*, propositions to accommodate this problem. Further details of this development are provided in the next Chapter that deals specifically with DNA evidence.

Chapter 6: Application of the principles of Case Assessment and Interpretation to specific evidence types

6.1 Cases involving DNA-analysis

The technique known as Multi-Locus Profiling (MLP) (Jeffreys *et al.* 1985) was the first DNA-profiling technique to be used for evidential purposes in criminal proceedings in the UK and was introduced to courts of law in the late 1980s. The output of the MLP technique was the familiar ‘bar-code’ pattern. Interpretation of the results was based on an estimate of the probability of sharing, by chance, the particular number of matching ‘bars’ that had been observed. Subsequent generations of DNA-profiling techniques involved analysis of specific loci, or sites, within the DNA molecule, and the results were in the form of electrophoretograms that showed which alleles (DNA types) were present at each of the loci. Interpretation of these results relies on databases of DNA-profiles from samples of people of the three main ethnic groups. From these databases, relative frequencies of occurrence for specific combinations of alleles can be calculated. These relative frequencies are then used to assign a probability (likelihood) for obtaining the match if the DNA had originated from someone other than, and unrelated to, the suspect with whom a match had been obtained. Evaluation of the weight of the evidence, in the form of a likelihood ratio, for the matching profiles can consist, in its simplest form, of the reciprocal of the match probability. The ensuing likelihood ratios for full profiles obtained using the SGM Plus® technique are then of the order 1 billion (Evetts *et al.* 2000b). In Bayesian terms, this means that the posterior odds of the DNA having come from the suspect would be equal to the prior odds of this event multiplied by 1 billion. Numbers such as 1 billion suggest huge evidential power and therein lies the beguiling attraction of

DNA-profiling evidence. However, in some situations, this power can be deceptive, and the attraction misleading, because of the influence of two key considerations – relevance and attribution.

At one end of the spectrum of cases, the body fluid that has been found and tested is accepted as directly relevant to the offence, and the DNA-profile that has been obtained can be attributed with certainty to that body fluid. In these cases, if the suspect denies any involvement whatsoever with the offence, then the power of the matching DNA-profiles can be translated directly to consideration of whether or not the suspect committed the offence – the value of 1 billion for the likelihood ratio provided by the DNA evidence would be appropriate also for the likelihood ratio at activity level and possibly also at offence level. An illustration of such a situation is provided by the following hypothetical case.

A victim of an assault alleges he was kicked several times to the head by one person. As a result of his injuries, the victim bled heavily. A suspect was arrested a short time after the incident and denies any involvement whatsoever with the attack. In particular, he denies being anywhere near the location of the offence and he has no explanation as to why he may have any blood on his clothing. On examination of his clothing and footwear, a single bloodstain is found on his right shoe. A DNA-profile obtained from this stain is a single, full profile and matches that of the victim. Non-bloodstained areas from around this single bloodstain are also tested but give no profiles whatsoever. It is therefore a safe assumption that the profile that has been obtained can be attributed to the bloodstain and not to any underlying DNA from other body fluids or tissues.

Evaluation of the weight of the scientific findings at activity level for this case would consider the probability of the findings under the following competing propositions:

- the suspect is the person who kicked the victim
- the suspect has nothing at all to do with the assault

The evaluation has, broadly, two components: 1) the probability of obtaining bloodstaining of this appearance on the shoe, and 2) the probability of obtaining matching DNA-profiles. If likelihood ratios were to be assessed for each of these two components, the value provided by the presence/appearance of the bloodstaining would be small compared to that provided by the matching profiles: the first likelihood ratio would be typically of an order between 10 and 100, while the second would be typically of the order 1 billion. Whatever the actual weight at activity-level that is provided by the first likelihood ratio, it would be 'swamped' by the weight of the second likelihood ratio. Evaluation of the weight of the scientific evidence at activity-level in this case is driven by the match probability for the DNA profile.

At the other end of the spectrum, although matching DNA-profiles may be obtained, they provide very poor likelihood ratios at activity-level. Another case, based loosely on the circumstances of an actual murder, will be used to illustrate this type of situation.

A witness sees a man, who was loading bags into the boot of a car, being approached by a second man. The witness describes this second man as wearing a hooded top which prevents the witness from seeing his face. When the hooded man is about 2 metres away from the man by the boot, he pulls a handgun from his pocket, takes aim and fires a single shot at the head of the man by the boot. The hooded man runs away and the man by the car falls to

the ground. The witness runs over to attend to the man on the ground but cannot resuscitate him. The man dies as a result of the gunshot.

On examination of the scene, the only item of significance that is found is a spent bullet case, believed to have been ejected automatically from the gun.

DNA-profiling tests on this bullet case give a weak, single DNA-profile. Tests to indicate the body fluid or tissue from which the profile has been obtained prove inconclusive. After several weeks, the police arrest a suspect and subsequently find that his DNA-profile corresponds with the profile from the bullet case. He does not have a gun in his possession and the gun that was actually used is never recovered.

In this situation, the weight of the DNA evidence at activity level will not necessarily be driven by the match probability (typically 1 in 1 billion) because, crucially, the relevance of the DNA-profile to the actual activity of firing the gun at the time of the offence is uncertain. Nor is it possible to attribute the profile to a specified body fluid/tissue. The only way to make progress in this situation is to try to make an assessment of the probability of the DNA-profile actually coming from the firer of the gun, whoever that was. The scientist could help with providing data, if it is available, on the occurrence of firers' profiles on spent bullet cases but there would also be an element of the trier of fact having a view on that probability, given the case circumstances. If that probability were to be assessed, by whomever, as 1, then the match probability would dominate the weight of the evidence. More likely, that probability would be assigned a value less than 1 and the resulting overall weight of the DNA evidence would be weaker than that provided solely by the match probability. The evaluation of the weight of the evidence becomes even more problematic if the suspect, while denying the shooting, admits to handling, in some

unspecified way, ammunition in the recent past. This information conditions the alternative proposition of the likelihood ratio and has a crucial bearing on the scientist's assessment of the probability of obtaining the matching DNA-profile given this alternative were true. It may be that the likelihood ratio at activity level, in this situation, would not be large and therefore of low probative value regardless of the very small match probability.

The paper by Evett *et al.* (2002), Appendix 4.1, demonstrated that, in these situations of uncertainty over attribution and relevance of DNA evidence, the power of matching DNA-profiles may well be greatly moderated. The authors applied the notion of the hierarchy of propositions to such cases and utilised the novel idea of *sub-source* level propositions as described by Evett *et al.* (2000a). This new concept clarified the interpretation of DNA evidence and demonstrated crucially that misleading opinion could be presented to the courts if the scientist operated solely at sub-source level instead of the more useful, more effective, activity level. The paper provided the logical structure to help practitioners deliver safer opinions.

6.2 Illicit drugs

The paper by Booth *et al.* (2002), Appendix 4.2, proposed an application of the CAI model to cases of alleged supplying of illicit drugs. Most drugs cases involve relatively straightforward identification and quantification of drugs, requiring little in the way of interpretation. However, a small number of cases involve more complex issues of dealing in illicit drugs. For example, there are cases in which the issue is whether or not a suspected 'dealer' has supplied a quantity of drugs to a particular 'user' from whom the drugs have been seized.

Booth *et al.* (2002) emphasised the importance of precision in the words used by the scientist in such cases and, for the first time, highlighted the importance of specifying the stage of the drugs 'supply chain' that the scientist's examination was addressing. If 'matches' between the suspected 'dealer's' sample and the 'user's' sample were found, then the traditional approach was to report the along the lines of:

... powders are likely to have come from the same batch ...

... the packaging is similar ...

The authors posed the question (page 123) of whether this phraseology provided maximum value, clarity and balance to the Criminal Justice System and, if not, what would constitute a better way?

The paper demonstrated how CAI principles could be applied to these cases and suggested that the specification of propositions and alternatives that were directly relevant to the issues in the case was central in encouraging robust, logical assessment and interpretation. This approach facilitated, for the first time, the proper, joint evaluation of the different types of scientific evidence that may be involved. Prior to this paper, the 'match' in appearance and chemical composition between the 'dealer's' and 'user's' drugs would be reported separately from the 'match' in appearance and composition between the wrapping/packaging of those drugs. The paper showed that, by considering appropriate propositions and alternatives at activity level, the joint weight of these 'matches' could be appraised. Furthermore, in adopting this approach, the scientist would gain insights into which datasets were appropriate to inform probabilities for the scientific findings.

The authors believed that application of the CAI model to this type of case provided four benefits – consistency, robustness, balance and added-value - over the ‘traditional’ approach.

The current author was the lead member of the team responsible for the approach described in this paper and provided the intellectual and philosophical framework to facilitate and develop, with his co-author specialist colleagues, application of the CAI model to drugs cases.

Chapter 7: Case Assessment and Interpretation and the construction of relevant databases

The United States' National Academy of Sciences (NAS), in a report on the state of forensic science (2009), concluded that (p184-185):

“There is a critical need in most fields of forensic science to raise the standards for reporting and testifying about the results of investigations. For example, many terms are used by forensic examiners in reports and in court testimony to describe findings, conclusions, and the degrees of association between evidentiary material (e.g., hairs, fingerprints, fibers) and particular people or objects. Such terms include but are not limited to “match,” “consistent with,” “identical,” “similar in all respects tested,” and “cannot be excluded as the source of.” The use of such terms can have a profound effect on how the trier of fact in a criminal or civil matter perceives and evaluates evidence. Yet the forensic science disciplines have not reached agreement or consensus on the precise meaning of any of these terms. Although some disciplines have developed vocabulary and scales to be used in reporting results, they have not become standard practice. This imprecision in vocabulary stems in part from the paucity of research in forensic science and the corresponding limitations in interpreting the results of forensic analyses.”

This comment by the NAS on the ‘*limitations in interpreting the results*’ begs the question of how should scientists interpret their findings. In the current author’s experience, scientists have interpreted their findings in a combination of ways ranging from a purely personal, intuitive approach that relies on experience, through to a demonstrably rational approach using published and verified data. The actual

approach taken will depend on the scientist's training, personal preference, corporate guidelines and the availability of reliable data. This variability in approach to interpretation is reflected in the variability of terminology in reports and statements to which the NAS report refers. If phraseology such as '*...could have come from...*', '*...cannot exclude...*' or, more simply, '*...match...*' are to be used, then the scientist requires knowledge primarily, and solely, about the variability of the material in question. Assessment of the weight of the evidence, if addressed at all, would revolve around an estimate of a frequency of occurrence. A key question would then be – what type of data should be collected to provide appropriate, useful frequencies of occurrence? On the other hand, if phraseology such as '*...likely to have broken the window...*' or '*...there is a high probability that the wearer of the clothing sat in the driver's seat of the car...*', then something more than simply knowledge of frequencies of occurrence is required. The scientist requires knowledge and understanding of the transfer and persistence of the material in question.

The profession of forensic science has been conscious, since early days, of the need to collect relevant data and efforts have been made over the years to do so. Broadly speaking, two types of data have been compiled:

- data to inform estimates of frequencies of occurrence of evidential materials;
- data to inform investigative aspects, e.g. the distribution of material after a particular activity; the maximum persistence of material after transfer.

In respect of the former data, the type of samples collected would generally be those that could be obtained conveniently. For example, glass objects that had been broken in the scientists' homes or environment would be collected to form a dataset on the frequencies of occurrence of different glass types in the general environment; fibre

samples would be collected from clothing items that were being examined within the laboratory to give an indication of commonness of different fibre types; footwear impressions were taken from footwear examined in the laboratory to give a frequency of footwear types. However, there were always discussions and challenges as to whether these were the most appropriate samples to collect. For example, do broken glass objects in households provide a good estimate of the frequency of occurrence of the types of broken glass on suspects' clothing?

In respect to the second type of data mentioned (those relating to distribution and maximum persistence), there is a question as to whether they are the most appropriate data to assist in the interpretation of evidential weight.

While ground-breaking work on the use of data to assess evidential weight had been published for some evidential materials, e.g. glass (Curran, Hicks and Buckleton 2000), the paper by Champod, Evett and Jackson (2004; Appendix 5.1) provided, for the first time, generic guidance on the collection of relevant data. The authors concentrated on source-level issues and developed a general treatment to inform the probabilities of a likelihood ratio. The key to the approach was the teasing out of two aspects of scientific evidence – observations that relate to the reference sample (of known origin) and observations that relate to the recovered sample (whose origin is unknown). The work showed that, provided certain assumptions were made, then, in principle, the scientist required data to inform three different types of relative frequency. An example involving footwear marks serves to illustrate the principle. It should be noted that the notation used in this example has been simplified from that used in the paper.

Assume a footwear mark has been found at the scene of a crime and that a suspect has been arrested on suspicion of committing the offence. His shoes and the scene mark are submitted for examination and, subsequently, a ‘match’ is found between the mark at the scene and the suspect’s shoe.

The likelihood ratio [from Equation 2] provided by this match is given by:

$$LR = \frac{\Pr(E | H_P, I)}{\Pr(E | H_D, I)} \quad [\text{Equation 3}]$$

where ‘E’ denotes the ‘match’, ‘H_p’ represents the prosecution proposition ‘the suspect’s shoe made the mark at the scene’, and H_D’ represents the defence proposition ‘the suspect’s shoe did not make the mark at the scene’.

Champod, Evett and Jackson (2004) decomposed ‘E’ into two parts:

Let ‘E_C’ represent the observations made about the mark at the scene

Let ‘E_D’ represent the observations made about test marks made from the suspect’s shoe

The likelihood ratio [Equation 3] becomes:

$$\frac{\Pr(E | H_P, I)}{\Pr(E | H_D, I)} = \frac{\Pr(E_C | H_P, I)}{\Pr(E_C | H_D, I)} \cdot \frac{\Pr(E_D | H_P, E_C, I)}{\Pr(E_D | H_D, E_C, I)} \quad [\text{Equation 4}]$$

If some simplifying assumptions are made, then this becomes:

$$\frac{\Pr(E | H_P, I)}{\Pr(E | H_D, I)} = \frac{1}{\Pr(E_C | H_D, I)} \cdot \frac{\Pr(E_D | H_P, I)}{\Pr(E_D | H_D, I)} \quad [\text{Equation 5}]$$

This analysis reveals that three sets of data are required to inform probabilities for the findings:

- footwear from people who have committed offences like the one in question (known as an ‘offender-related’ database);

- footwear from people, like the suspect, who could be suspected of the offence but who are, in fact, innocent (a so-called ‘innocent suspects’ database)
- footwear marks found at the scene of similar crimes (a ‘crime-related’ database).

It may be that in some circumstances, and for certain evidence types, these sets would provide very similar estimates of the relative frequencies for the matching mark. If that were so, then the probabilities for the observations would be very similar and the right hand side of the extended likelihood ratio [Equation 5] would be equal to 1. The value of the ‘match’ is then given by the reciprocal of the frequency of occurrence of the matching mark at scenes of similar crimes. In these particular circumstances, therefore, only one data set is required to evaluate the likelihood ratio. However, this would only be known if the scientist had gone through the process of identifying these data sets and checking if they provided significantly different estimates of the relative frequencies.

Champod, Jackson and Evett (2004) provided, for the first time, clear guidance on the collection and construction of appropriate data sets to help scientists form and express more robust, reliable opinions.

Chapter 8: A unifying view of the role of forensic science

Over the course of a large number of workshops and other opportunities with experienced practitioners, the Case Assessment and Interpretation (CAI) model was applied by the CAI development team (which included the current author) to the full range of main-stream forensic science disciplines from 1998 onwards. That experience led to a deeper understanding of the model's suitability and its benefits and limitations. It also led to a novel view, and a formalisation, of the role of forensic science. From that, new insights were gained into the skills and knowledge required to perform the different aspects of that role. These developments were reported in a set of three papers that are the subject of this Chapter.

A paper by the current author (Jackson 2000; Appendix 6.1) reviewed the contribution that forensic science had made over the preceding sixty or so years, focussing particularly on how scientists had interpreted their findings and communicated their opinions. The limitations of various words and phrases used conventionally by scientists to express their conclusions were explored. A set of guiding principles was offered (page 85) that, in Jackson's opinion, provided the framework to develop and improve the delivery of forensic science services:

- add value
- act impartially
- be transparent
- provide sustainable, robust evidence and be expert

This framework was based on application of Bayesian inference and, in particular, on appraisal of a likelihood ratio as a logical means of assessing and presenting the significance of scientific findings.

Several years later, this set of guiding principles was reflected in the standard procedure for offering evaluative opinion written by the Association of Forensic Science Providers (2009, page 161):

- balance
- logic
- transparency
- robustness

Focussing for the moment on the principle of ‘balance’, the original version of the Case Assessment and Interpretation model required the scientist to define a pair of mutually exclusive propositions relevant to the case in question. For some practitioners and for some cases, this was a relatively straightforward task, particularly if the propositions were centred on source or sub-source level. However, there were many cases in which it was difficult to define propositions. Part of this difficulty seemed to be the inability of practitioners to understand what they were trying to achieve with their examinations. In a subsequent paper (Jackson *et al.* 2006; Appendix 6.2), the author of this thesis provided a novel solution to the problem by suggesting that, prior to framing *propositions*, practitioners should define the *issues* in the case from the viewpoint of the customers – the police, prosecutors or defence. The introduction of this step helps the scientist to focus on the key issues in the case and helps therefore to direct the purpose of the examination.

The following example, based loosely on an actual case of murder, may help to illustrate this concept:

The scene of the crime is the home of an elderly married couple who had suffered injuries that had bled profusely leading to their death. A bloodstained towel has been found at the scene of a murder and, from other evidence, it is believed that the towel belonged to the house and had been used by the offender to wipe up blood at the scene. The police submit the towel, together with relevant reference samples from the deceased and a suspect, with a request for DNA analysis.

On first glance, it may seem obvious and straightforward in this case to develop a pair of sub-source propositions along the lines of:

- Some of the DNA on the towel came from the suspect
- The DNA came from people other than, and unrelated to, the suspect

If this were so, then interpretation of the results of the analysis would also be relatively straightforward, requiring nothing more than an assessment of match probabilities for the likelihood ratio. However, if the scientist were to probe more deeply into the case, it may be that, through discussions with the customer, the case circumstances reveal that the suspect denies he is the offender but says he was in the house one week earlier and had handled the towel then. Clearly, the issue then is not about the origin of the DNA (sub-source level) but about whether the suspect handled the towel at the time of the offence or had handled it one week earlier (activity level). With this information, and following the guidance in the paper by Jackson *et al.* (2006), the scientist would proceed by specifying an appropriate, relevant issue at activity level such as:

- ‘Did the suspect handle the towel at the time of the offence rather than one week earlier?’

Once the issue has been clarified, specified and agreed with the client, a pair of mutually exclusive propositions based on this activity-level issue can be developed:

- The suspect handled the towel at the time of the offence;
- Someone else handled the towel at this time; the suspect handled it one week earlier.

Interpretation of the findings would then depend on consideration of the quantity, distribution and presence of mixtures of DNA-profiles, not solely on frequencies of occurrence or match probabilities, as would be the case if the scientist stayed at sub-source level. Furthermore, if opinion were to be presented in this case at sub-source level, then the jury may well be misled into thinking that the DNA evidence was very powerful and compelling evidence for the guilt of the suspect. This concern, of vital significance for the effective administration of justice, was explored earlier in this thesis (Chapter 6.1).

A further, important contribution of the paper by Jackson *et al.* (2006) was to take forward the development of the distinction between the role of a forensic scientist as an ‘investigator’ and as an ‘evaluator’, a notion first aired by Evett *et al.* (2000). As lead author, Jackson suggested that ‘investigators’ provide opinions that could be classified either as 1) *explanations* generated to explain the results of their examinations or 2) *posterior probabilities* of propositions, after the results of the examination are known. In contrast, ‘evaluators’ provide *likelihood ratios* for the results of the tests or observations. For example, when the scientist is involved in the early stages of a case, such as at the scene of a crime, then they will be offering opinions such as - ‘*Possible sources of the black fibres recovered from the victim’s body include carpets and upholstery*’ (explanations), and ‘*The seat of the fire is very*

likely to be the armchair in the far right-hand corner of the living-room' (a posterior probability). When the case progresses and a suspect has been arrested, then the opinions given by the scientist would be an expression of a likelihood ratio, and would be phrased along the following lines – '*The findings are far more likely to be obtained if the victim had been transported in the suspect's vehicle rather than in some other, unknown vehicle*', or '*The findings provide moderate support for the view it was the suspect, rather than someone else, who poured petrol over the armchair*'.

With the aim of understanding more deeply the nature of forensic science opinions, the paper by Jackson (2009; Appendix 6.3) offered a review and an exploration of the wide range of expressions commonly used by forensic scientists. A categorisation of these expressions following the 'investigative' or 'evaluative' structure (Jackson *et al.* 2006) was presented and it was suggested that, if an opinion can be so classified, then the skills and knowledge required to give such an opinion can be defined.

For *explanations*, the skills and knowledge would include the ability to

- 1) generate a full range of realistic hypotheses to account for the findings;
- 2) communicate the strengths and, more importantly, the limitations of such opinion.

For *posterior probabilities*, the scientist should have the ability to

- 1) generate all relevant prior hypotheses;
- 2) assign realistic, fair prior probabilities for these hypotheses;
- 3) assign justifiable, robust likelihoods for the findings;
- 4) compute posterior probabilities;

5) communicate clearly not only the opinion in the form of a posterior probability but also the justification of such an opinion.

For *likelihood ratios*, the requirements would include:

- 1) the ability to define the issues that are relevant to the case;
- 2) the ability to translate these issues into appropriate propositions and alternatives based on the case circumstances;
- 3) the knowledge to help assign robust likelihoods for the findings;
- 4) the skills to explain what an opinion in the form of a likelihood ratio actually means and how such an opinion can be used by the triers of fact.

Not all of the words and phrases used by scientists fell easily into this structure, primarily because the words or phrases were of obscure or confused meaning. The recommendation in the paper was that use of these expressions in scientific language should be avoided.

This categorisation of opinions, together with the description of the skills and knowledge required to give those opinions, provided practitioners with a new way of checking the robustness of their opinions. It also gave the people who use expert services, i.e. police and lawyers, a framework they could use to challenge forensic science evidence and to check whether the expert's opinion was justifiable.

This was a novel view of the role of forensic scientists and of the nature of the opinions that they provided. It helped clarify, for both providers and consumers, the

contribution that forensic science makes to investigations and to the criminal justice process.

It is difficult to know precisely what influence the concepts described within the three papers of this Chapter (and papers from other Chapters) had on the Association of Forensic Science Providers (AFSP) when they developed their standards for the formulation of evaluative opinion (Association of Forensic Science Providers 2009). However, the AFSP do make reference to at least one of the papers and there is much resonance in the language of the AFSP document with that of the papers.

Chapter 9: The current status of Case Assessment and Interpretation in Criminal Justice Systems.

The philosophical approach that underpins the process of Case Assessment and Interpretation is one of inductive, Bayesian inference. Generally, inference can be viewed as a very personal activity that depends on the specific thought-processes of the individual making the inference. It can be argued that, by applying the logical structure of inductive Bayesian inference contained within CAI, there is no room for individual variation in approach by scientists. However, there are a number of key steps within CAI that are subject to decisions that can only be made by the scientist using judgement. Some of these decisions can be influenced and controlled within the scientist's domain. For example, decisions on assigning probabilities for scientific findings will depend on the scientist's own knowledge, understanding and experience of the evidence type. However, other decisions will depend on factors outside the control and possible knowledge of the scientist. For example, definition of the issues to be addressed must be influenced by knowledge of the case circumstances but the scientist's awareness of the circumstances will necessarily be limited. Furthermore, decisions on which items to select for examination may depend not only on the predicted evidential value but also on the constraints of the customer's budget and deadlines if, as in England and Wales, there is a formal, commercial relationship between suppliers and customers. Even if there is no such relationship, as in Scotland and the Republic of Ireland, there is still the pressure on suppliers of forensic science services to use their resources efficiently and economically.

Jackson and Jones (2009; Appendix 7.1) argue that effective application of CAI relies not only on the competence and expertise of the scientist but also on the active

engagement and participation of customers, in the widest sense, with the scientists providing the service. The overall aim of the CAI model was that of delivering a service meeting the needs of the Criminal Justice System. These needs can only be met if the needs are properly understood – and this can only come from effective engagement of providers with customers. The authors argue that there are various constraints and barriers to such engagement and that these include:

- poor communication, leading to misunderstandings between providers and users
- insufficient time to assess needs
- limited budgets to purchase forensic science service
- inefficient or ineffective working practices in both provider and customer organisations
- weak challenge and counter-balance to prosecution-commissioned forensic science
- customers' focus on purchasing 'commodities' rather than 'value-added' services
- the adversarial nature of the criminal justice system

The current commercial arrangements between police and forensic science providers in England and Wales, summarised in the National Forensic Framework Agreement (National Policing Improvement Agency 2010), are described by Lawless (2010; page 16) as - *'... tight controls surrounding the way in which forensic services are provided'*. Lawless continues: *"The notion of 'partnership' found in the CAI literature is significantly under-represented here"*.

While the specific details of the National Agreement are somewhat difficult to obtain, anecdotal information suggests that, within this Agreement, the police can commission ‘source only’ products. If this were in fact true, then to do so runs the risk, at best, of wasting money and resources or, at worst, of providing misleading evidence.

Jackson and Jones (2009, page 11) argue that effective application of the principles of Case Assessment and Interpretation not only requires competent individual scientists and customers but also requires each organisation to have in place ‘supportive procedures’ that facilitate the process. Such procedures include excellent communication and common understanding between all the actors in the process, quality standards that assure competence, and the provision by supplier organisations of data to assist the scientist. It remains to be seen, given the commercial pressures on suppliers to keep costs of ‘overheads’ such as research low, how suppliers will be able to meet this latter requirement of generating and providing useful data.

Finally, the authors provide a step-by-step checklist to guide scientists, and others, through the process of applying the principles of Case Assessment and Interpretation (CAI) in individual cases. Cook *et al.* (1998a) provided the original description of the CAI model, but the version in Jackson and Jones (2009) incorporated the subsequent developments and improvements of the hierarchy of issues and the classification of investigative and evaluative opinions. This approach provided, for the first time, a logical, unifying process for all forensic science practitioners.

Chapter 10: Summary and review

10.1 Introduction

It has been argued in this thesis that applying the Case Assessment and Interpretation model, with its underlying requirement to assess likelihood ratios, to the examination of cases helps the forensic scientist, firstly, to make effective decisions on which items to examine and, secondly, to provide robust, balanced interpretations of the results of those examinations. This approach is set within the accompanying concepts of the ‘hierarchy of issues’ and a Bayesian classification of expert opinions. Numerous examples of how the CAI model can be applied to casework are provided in the papers appended to this thesis and the reader is directed to these examples for illustration of the underlying concepts and the value of the approach. In particular, the case example – ‘A Simple Transfer Case’ – presented in the paper by Jackson and Jones (2009; Appendix 7.1) walks the reader through the key stages of the CAI model and illustrates both qualitative and quantitative aspects of the model.

While the case studies provided in the appended papers illustrate the application and the benefit of the approach, further demonstration of the usefulness of CAI can be achieved through revisiting the cases described in Chapter 2 and assessing the impact that CAI may have had in those cases. At the end of the Chapter, there is a discussion of two more recent cases in which consideration of a likelihood ratio did have a significant impact on the eventual outcome of the case.

The process of Case Assessment and Interpretation can be summarised by the following steps, based on the ‘checklist’ provided by Jackson and Jones (2009):

1. Define the customer requirement

- the provider makes efforts to understand the case circumstances, the allegations and the uncertainties.
2. Assess how forensic science can help
 - the provider assesses the most likely outcomes, and their potential evidential value (likelihood ratio) or investigative value, for the different examinations that could be undertaken and advises the customer accordingly.
 3. Agree on a case examination strategy
 - following informed conversations between the customer and provider, an agreement is reached on which items to examine with which tests.
 4. Carry out the agreed strategy
 5. Interpret the results
 - if evaluative, consider and refine likelihood ratios
 - if investigative, consider explanations or posterior probabilities
 6. Communicate
 - convey clearly the conclusion as it relates to the original purpose of the examination

With the historic cases to be revisited here, it is impossible to carry out all the steps in this process. The main focus, instead, will be a consideration of whether the issues were investigative or evaluative and, if the latter, what would have been appropriate pairs of propositions and what weight of evidence might have been provided?

10.2 The case of John (Jack) Preece

In the discussion of this case in Chapter 2 (see 2.2.1.), several questions were raised about how the scientist could have proceeded. One of these questions was ‘... exactly how should experts give evidence in an *unbiased manner*?’ How might the CAI model have helped in this respect?

In the first instance, the CAI model would have directed the scientist to understand the issues in the case and, from these, construct pairs of propositions that reflected the respective positions of the prosecution and defence. For greatest value to the court, it has been proposed that these propositions should be at activity-level. However, given the lack of detail of the specific circumstances of this particular case, this thesis will retreat to source-level and suggest a pair of propositions that is generic to many cases of this type:

H_P - the semen came from Mr Preece

H_D - the semen came from someone other than Mr Preece

(where H_P represents the prosecution allegation and H_D the alternative argument)

If this pair of propositions had been adopted in the case, then the next step, following the CAI model, would have been to evaluate a likelihood ratio for the findings in relation to ‘grouping’ tests. The key issue at the trial was whether or not any group over and above that of the female victim’s own group had been detected in the mixed seminal and vaginal staining from the victim. The original, prosecution-commissioned expert in the case asserted that he had detected reactions for group A that were not attributable to the victim herself, who was also of group A, but could be attributed to the semen present in the staining. Other experts giving evidence at the Appeal hearing had the opposite view – no group attributable to the offender (whoever that

may have been) could be discerned from the results. Instead of engaging in an argument about which expert opinion is correct, or which expert is to be believed, a probabilistic approach to the evaluation of the findings could have proceeded in the following way.

Let 'A' represent the detection of group A.

Let 'I' represent all the conditioning background information influencing assignment of probabilities.

The weight of evidence provided by the detection of group A would be given by the likelihood ratio (from Equation 2):

$$LR = \frac{\Pr(A | H_P, I)}{\Pr(A | H_D, I)} \quad [\text{Equation 6}]$$

However, the uncertainty over whether or not the group A that was detected actually came from the offender (whoever that may have been) has to be accommodated in this formulation of the likelihood ratio. This uncertainty can be modelled in the following way:

Let 'O' represent the event of an offender's group (whatever that was) being detected in the mixed staining.

Let 'NO' represent the event of an offender's group not being detected in the mixed staining. The values of the probabilities of these two events would be complementary and would have been informed by factors such as the relative quantities of semen and vaginal material present in the mixture, together with the degree of expert knowledge on the detection of groups in such mixtures.

The likelihood ratio [Equation 6] can be extended as follows:

$$\frac{\{\Pr[\text{NO}|\text{H}_p, \text{I}] \times \Pr [\text{A}|\text{H}_p, \text{NO}, \text{I}]\} + \{\Pr[\text{O}|\text{H}_p, \text{I}] \times \Pr [\text{A}|\text{H}_p, \text{O}, \text{I}]\}}{\{\Pr[\text{NO}|\text{H}_D, \text{I}] \times \Pr [\text{A}|\text{H}_D, \text{NO}, \text{I}]\} + \{\Pr[\text{O}|\text{H}_D, \text{I}] \times \Pr [\text{A}|\text{H}_D, \text{O}, \text{I}]\}}$$

[Equation 7]

On the assumption that Mr Preece was no different from other potential offenders in his propensity to having his group detected, the term $\Pr[\text{NO}|\text{H}_p, \text{I}]$ in the numerator would be equal to the equivalent term in the denominator, $\Pr[\text{NO}|\text{H}_D, \text{I}]$. The probability can therefore be written as $\Pr[\text{NO}|\text{I}]$ and constitutes the critical issue that was argued at the Court of Appeal – how likely was it that the group of the offender, whoever he may have been, was discernible in the results? The decision made by the Court on this issue, based on the expert evidence that they heard, was that it was certain (probability 1) that no offender group was detected. However, let us assume maximum uncertainty for the probability $\Pr[\text{NO}|\text{I}]$ and assign a value of 0.5. As the events ‘O’ and ‘NO’ are mutually exclusive and exhaustive, it follows that $\Pr[\text{O}|\text{I}]$ must also have a value of 0.5.

The term $\Pr [\text{A}|\text{H}_p, \text{NO}, \text{I}]$ can be assigned a value of 1 because, on the conditioning that the offender’s group had not been detected and that the victim is of group A, it is practically certain, barring experimental and human error, that group A would have been detected. If the offender’s group had been detected, and that offender was Mr Preece, then it is practically certain that group A would have been detected, given Mr Preece was of group A. The term $[\text{A}|\text{H}_p, \text{O}, \text{I}]$ could then be assigned a value of 1.

Turning now to the denominator, a value of 0.5 could be assigned for the terms $\Pr[\text{NO}|\text{H}_D, \text{I}]$ and $\Pr[\text{O}|\text{H}_D, \text{I}]$, following the argument earlier. If no offender’s group had been detected, and because the victim was group A, it would be practically certain

that group A would have been detected: the term $\Pr[A|H_D, NO, I]$ would equal 1. If the offender's group had in fact been detected, but this offender was not Mr Preece, then the term $\Pr[A|H_D, O, I]$ would be a function of the occurrence of group A in a relevant population of people who could be suspected as being the source of the semen. On the assumption that the UK Caucasian population was the relevant population, then a probability of 0.4 could be assigned for this term, based on the approximate proportion of group A in this population.

Substituting the values as described for the probabilities, the likelihood ratio [Equation 7] becomes:

$$LR = \frac{\{0.5 \times 1\} + \{0.5 \times 1\}}{\{0.5 \times 1\} + \{0.5 \times 0.4\}}$$

[Equation 8]

which reduces to:

$$LR = \frac{1}{0.7}$$

[Equation 9]

to give a likelihood ratio of approximately 1.4.

The effect of this magnitude of likelihood ratio would be to multiply the prior odds of the semen having come from Mr Preece by a factor of approximately 1.4. Following the current verbal convention for expressing the likelihood ratio as a strength of support (Evetts 2006), this would be described as limited (or weak) support for the prosecution proposition over the defence proposition. Whichever way the grouping results are viewed, most would agree that, given a likelihood ratio of 1.4, the evidence is of very limited value and should have been reported as such by the original expert.

10.3 The case of Judith Ward

The question was posed in the discussion of this case (see 2.2.2) whether, at the time of the original examination, there was a more acceptable, more logical and robust guiding structure for forming forensic science opinion. The answer to the question is yes, there was a logical structure available, but it was not widely known among, and certainly not taught to, main-stream forensic science practitioners. Had it been known, how might the experts have improved their interpretation in this case?

The first improvement would have been in the phrasing of opinions. Wide variation in the way opinions were expressed was apparent in the judgment¹⁶. Examples included:

- *'... consistent with the opinion that contact of the hands with an explosive substance could have occurred'*
- *'... consistent with the opinion that contact of the inside of the bag and commercial explosive has probably occurred'*
- *'... results showed that Miss Ward had probably been in contact with a commercial explosive'*
- *'... Miss Ward must have kneaded explosives.'*

The word 'contact' has a multitude of meanings and could be described as 'pseudo-activity' level (Evetts *et al.* 2000a) – essentially, the evaluation of the weight of evidence would follow a source-level appraisal while giving an appearance, falsely, of appraisal at activity-level. The CAI model encourages the scientist to phrase issues and propositions at activity-level in order to focus, and make clear, the basis of their

¹⁶ Ward (n 6).

interpretation. As with the previous case of Mr Preece, the precise circumstances and allegations in this case are not known to the current author and it is impossible to know what would have been more relevant phrasing, at ‘activity-level’, in this case. However, for illustration purposes only, some assumptions on the circumstances will be made and examples of improved phrasing for the issues in this case can be suggested on this basis. The following are just some examples of the issues that may have been relevant and appropriate to this case:

- whether or not Ms Ward planted the explosive device(s) (activity-level)
- whether or not Ms Ward kneaded explosives 60 hours before her hands were swabbed (activity-level based on the time of the last bombing incident)
- whether or not the traces on her hands came from the devices’ explosive (source-level)

Pairs of relevant propositions could be generated from these issues, taking into account the prosecution allegations and the defence assertions. Probabilities for the findings given the truth of these propositions, and given any conditioning information, could be assigned. The robustness of these probabilities would depend on the scientist’s knowledge and expertise but, as the current author does not have expert knowledge of explosives, no realistic probabilities can be offered here. However, based on evidence presented to the Court of Appeal¹⁷, it could be argued that the likelihood ratios would not have been too informative and, had the opinion been presented in the form of an expression of likelihood ratios, or in the form of investigative opinions with suitable riders about their limitations, then there would have been less risk of the jury at the original trial being misled.

¹⁷ *Ward* (n 6).

If the scientist had had available the classification of opinions (Jackson 2009), he would have had the framework to recognise that the opinion of “... *results showed that Miss Ward had probably been in contact with a commercial explosive*” was a posterior probability for a proposition. Posterior probabilities for propositions either should not be offered at all or, with the knowledge and permission of the court, should be offered with a rider about assumptions and limitations.

The opinion ‘... *Miss Ward must have kneaded explosives*’ would have been recognized by the scientist as an unjustifiable, categorical, posterior probability and would not have been given.

The case of Judith Ward provides a good example where a clearer distinction between investigative opinions and evaluative opinions, as proposed by Jackson *et al.* (2006), would have helped all concerned (scientists, police, lawyers, jury) to appraise the scientific results in a more balanced, robust way. There appeared to be traces of explosives on a number of different items from various sources relating to Judith Ward. In investigative mode, the scientist could have offered explanations for these traces but would have had to ensure that these explanations were not biased one way or the other and that they were exhaustive as far as was feasible. It would have been necessary for the scientist to have stressed the list of explanations was not truly exhaustive and to have pointed out that no opinion was being offered on the probability of explanations that had been put forward.

10.4 The case of the ‘Birmingham Six’¹⁸

This case raised again the question of how the expert could have presented his findings in a more reliable, more robust, more acceptable way (see 2.2.3). The scientist was reported as giving an opinion that he was ninety-nine percent certain that Mr Power and Mr Hill had been in contact with commercial explosives (Mullin C. 1990). Following the models provided by Evett *et al.* (2000b) and Jackson *et al.* (2006), this could be classified as an ‘investigative’ opinion in the form of a posterior probability at ‘pseudo-activity’ level. Putting aside the problem with the word ‘contact’, the main concerns with opinions expressed as posterior probabilities are opacity and possible bias - opacity because the prior probabilities have not been exposed; possible bias because the prior probabilities could be unjustifiably weighted towards the prosecution (or defence) proposition.

As with the Judith Ward case mentioned earlier, the precise circumstances and allegations in the Birmingham Six case are not known to the current author and it is impossible to know what would have been more relevant, ‘activity-level’ phrasing of propositions in this case. However, again for illustration purposes only, some assumptions on the circumstances will be made and examples of improved phrasing for the issues in this case can be suggested on this basis. The following are just some examples of the issues that may have been relevant and appropriate to this case:

- whether or not Mr Power planted one of the explosive devices (activity-level)
- whether or not Mr Power kneaded explosives approximately 20 hours before his hands were swabbed (activity-level based on an assumption on the time interval between assembling the device and swabbing)

¹⁸ *McIlkenny* (n 12).

- whether or not the traces on his hands came from the device's explosive (source-level)

Again, as in the Judith Ward case, clarity between investigative opinion and evaluative opinion would have been helpful. If evaluative issues had been discussed and agreed prior to the trial, opinions based on appraisal of likelihood ratios could have been offered and the impact of conditioning information 'I' on the probabilities for the findings would have been recognised and accommodated. Opinions founded on this type of appraisal would have helped the court by providing more balanced, more robust evidence. Alternatively, in investigative mode, explanations for the findings could have been put forward, along with a clear warning that a list of explanations is not exhaustive nor is there any probability attached to the truth of the explanations. Included in that list should have been reasons why apparently positive reactions would have been obtained from hand swabs given the alternative proposition that the defendants had not handled explosive.

10.5 The case of Stefan Kiszko

It has been argued earlier (see 2.2.4) that the issue in this case was one of profoundly poor interpretation of the value of the scientific findings and that, if the scientist involved had had a robust interpretational model, then perhaps he would not have presented evidence that was used to convince the jury that an innocent man had committed a serious crime. It was suggested that, probably, the scientist had simply presented his finding of the presence of semen on the clothing, without mention of the significance of the absence of spermatozoa in Mr Kiszko's semen. It is perhaps trivial, given the current knowledge and awareness among most practitioners about

robust interpretation, to apply CAI thinking to this case but doing so will highlight situations in which categorical opinions may be given.

The issue in the case can be stated at ‘source-level’, as in the *Preece* case mentioned earlier:

H_P - the semen came from Mr Kiszko

H_D - the semen came from someone other than Mr Kiszko

If H_P were true, then the probability of finding semen containing spermatozoa would be zero, according to the medical and other evidence that was adduced at the Appeal¹⁹ and that was available at the original trial. Accordingly, whatever value would be assigned to the probability for these findings given the truth of the alternative proposition H_D , the likelihood ratio would always turn out to be zero, and the posterior probability for H_P would therefore be zero – the prosecution hypothesis is not true.

Applying CAI may seem an overly complicated way of dealing with the scientific evidence in this case - it is probably obvious to even lay-people that Mr Kiszko could not be the donor of the semen at the scene. However, application of a likelihood ratio approach can underpin all assessments of expert evidence and should be the bedrock of all experts’ thinking.

¹⁹ *Kiszko* (n 14) 10F.

10.6 The case of Barry George²⁰

Mr George was convicted in the year 2000 of the murder in 1999 of Jill Dando. One of the three main strands of the prosecution case at Mr George's trial²¹ was the presence of a single particle of firearms discharge residue (FDR) in a pocket of a coat belonging to Mr George. Mr George's first appeal against conviction failed. But a second appeal was allowed after his case had been reviewed by the Criminal Case Review Commission (CCRC). The CCRC had commissioned a report into the FDR evidence and, on the basis of the findings in that report, Mr George's case returned to the Court of Appeal.

Many of the opinions expressed by scientists at original trial about the FDR evidence can be classified, following the scheme proposed by Jackson (2004), as 'explanations' at source-level, e.g. '*...consistent with having come from the cartridge used in the killing*'.²² There was also considerable discussion at that trial about the probability of 'secondary contamination', i.e. the FDR having arisen not through the act of firing the gun but through transfer via an intermediary. Some of the phrases used to convey opinions about this issue could possibly be seen as 'posterior probabilities' at activity-level, e.g. '*...it was equally unlikely that it was the result of the appellant firing a gun a year before*'.²³ No record could be found at the original trial of an appraisal of a likelihood ratio for the FDR findings.

The Court at the second appeal focussed primarily on the interpretation and communication of the value of the firearms discharge residue. The Court appeared to be persuaded that the approach to assessing the weight of evidence provided by the

²⁰ *R v Barry George* [2007] EWCA Crim 2722.

²¹ *ibid* 46.

²² *ibid* 9.

²³ *ibid* 39.

FDR result was through consideration of a likelihood ratio at activity-level, as described in the CCRC report. The Court said ‘*The most significant (fresh) evidence that we have received owes its origin to the initiative of Dr Ian Evett ..*’²⁴, adding later:

*...Dr Evett was in the course of developing with colleagues a technique called Case Assessment and Interpretation (CAI). The object of this is to clarify before evidence is examined and analysed the likelihood of the examination achieving particular results on two different hypotheses or propositions. This technique facilitates the drawing of appropriate conclusions from the results actually obtained on the examination.*²⁵

The two propositions that Dr Evett said he and the FDR expert considered in discussions after the original trial were:

H_P - the appellant was the man who shot Jill Dando

H_D - the appellant was not the man who shot Jill Dando²⁶

The FDR expert went on to assign values of approximately 1 in 100 for the probability of the findings if either of the propositions were in fact true. As the numerator and denominator of the likelihood ratio are of the same broad value, the likelihood ratio for the findings would be equal to 1 and, as concluded in a Forensic Science Service (FSS) report commissioned by the CCRC, ‘*provides no assistance to anyone asked to judge which proposition is true*’.²⁷

²⁴ George (n 20) 14.

²⁵ ibid 16.

²⁶ ibid 17. Note that this pair of propositions has been developed from an ‘activity-level’ issue.

²⁷ ibid 23.

Had the FDR evidence been appraised along these lines originally, and if the case had gone before a jury, then the jury would have heard that the FDR evidence was essentially neutral, providing no assistance on whether it was Mr George or some other person who had fired the gun that killed Jill Dando.

The Court of Appeal ruled that the verdict was unsafe and that the conviction would be quashed. Mr George's case went to a retrial and, at that trial, the FDR results were not admitted as evidence despite the prosecution's attempt to adduced the evidence. The jury subsequently found Mr George not guilty of murder.

This case is a clear example of the strength and benefit of the 'I' part - Interpretation - of the 'CAI model'. Whether there would have been benefits from the 'A' part of 'CAI' - the Assessment phase – in this case is debatable. However, it can be seen from the expert's evidence to the Court of Appeal that the most likely outcome (probability of 0.99) of examination of the coat, given the truth of either proposition, would have been no particles. Again, as the numerator and denominator would be of equal value (0.99), the likelihood ratio for finding no particles would be of value 1 and therefore of no evidential weight. Therefore, had the expert pre-assessed examination of the coat for FDR, then he would have predicted that the most likely outcomes of the examinations were no or a few particles, with both outcomes being of no evidential weight. Whether or not the police or prosecution, given this assessment, would still have requested the examination is unknown. It may well have been that the police would indeed have requested the examination for investigative purposes and, if that were so, then the opinion from the expert should have been in the form of investigative opinions, i.e. a list of 'explanations'. This approach is recommended in

another FSS document ('The assessment, interpretation and reporting of firearms chemistry cases') that was quoted in the judgment:

... the FSS has adopted a cautious approach to reporting LOW levels of residue and no evidential value can be offered.

From an investigative point of view, LOW levels of residue may nonetheless have some value; for example, finding a LOW level on a discarded item such as a glove may give a significant lead to a police investigation. When an officer is given information on LOW levels in an investigative submission, he must be made aware that in most cases it is unlikely any evidential weight can be attached to the finding.²⁸

10.7 The case of R v T²⁹

Despite the Court of Appeal in the Barry George case appearing to endorse a 'likelihood ratio approach' to the evaluation of scientific evidence, the judgment handed down in the more recent case of *R v T* appears to have taken an opposite view.

The Court made numerous, mainly adverse, comments on the use of likelihood ratios and Bayes' Theorem. The key complaint of the Court would seem to revolve around the availability of relevant and reliable data to inform the probabilities of a likelihood ratio and, while the case involved comparison of footwear marks, the Court's ruling could be interpreted as applicable to other forms of expert evidence. The Court compared the availability of data in the field of DNA with that for footwear marks and commented that:

²⁸ *George* (n 20) 21

²⁹ *R v T* [2010] EWCA Crim 2439.

*The most in our judgment that can be derived from a comparison of pattern and size (of footwear marks) is that it can form part of the judgment on how common the pattern or size is. It is impossible to see on the present state of information how any mathematical figure can be properly calculated to express a more definitive evaluative opinion – there are far too many variables and uncertainties in the data.*³⁰

These comments reflect the problem of misunderstanding between, on the one hand, experts who are trying to apply logical inference when forming opinions and, on the other hand, the legal profession who have the difficult task of incorporating expert opinion into an overall view of the weight of the totality of the evidence. The Court in *R v T* appeared to view the evaluation of a likelihood ratio as a strictly mathematical process that can only be used where there are precise, sufficient data. It may be that the Court did not understand the nature of subjective probabilities, and their use in dealing with uncertainty, and that this may have been due to a failure on the part of the expert witnesses in this case to explain this concept adequately for the Court. While the Court did talk about ‘*how common a pattern or size is*’ in the previous quote, and also referred to the use of a ‘*statistical database*’ in glass evidence³¹, they fail to show how these data can be used logically and coherently to form an evaluation of the weight of the findings. The only guidance given, referring to glass evidence, was:

...an expert can give an opinion using a statistical database by simply using that database and expressing an opinion by reference to it, without recourse to

³⁰ *R v T* (n 29) 85.

³¹ *ibid* 91.

*the type of mathematical formula used in this case or to any form of Bayes' Theorem.*³²

This comment begs the question - how should the expert use the data? In that respect, the judgment is quite opaque.

It is also very difficult to understand why this judgment is quite dismissive of the positive comments by the same Court on the use of likelihood ratios by Dr Evett in the *Barry George* case:

*It is apparent ... that Dr Evett was using a likelihood ratio approach, as advocated by him in the papers to which we have referred. However, Lord Phillips CJ (in giving the judgment of the court) did not consider the merits of the approach in the judgment or how it was consistent with the views expressed in Adams, Doheny and Adams (No2). It appears simply to have formed part of the background to the court's consideration of the overall evidence in the case.*³³

This is in striking comparison with the comment in the *Barry George* judgment- '*The most significant evidence that we have received owes its origin to the initiative of Dr Ian Evett ...*'³⁴

Redmayne *et al.* (2011), Evett (2011) and Berger (2011) all provide critique and comment on the *R v T* judgment and all emphasise the logical, robust approach to the evaluation of the weight of expert evidence provided by an assessment of a likelihood

³² *R v T* (n 29) 91.

³³ *ibid* 90.

³⁴ *George* (n 24) 14.

ratio. It remains to be seen if the $R \vee T$ judgment will be an influential judgment on the admissibility of expert evidence, or whether subsequent judgments will present a more soundly-based argument.

10.8 Case assessment

So far in this Chapter, the emphasis has been on the Interpretational aspects of CAI; the Assessment phase of CAI has only briefly been mentioned. This reflects the emphasis that the Criminal Justice System currently has on outcomes of prosecutions and the central role that interpretation of expert findings plays in some cases. When the CAI model was first promulgated, it was in response partly to the requirement to provide value-for-money services in forensic science. The model facilitates the making of rational decision, prior to any substantive work being undertaken, in those cases in which there are choices to be made about the items to examine and the techniques to deploy.

In recent years, there has been a trend towards customers asking for forensic science services in a far more prescriptive way. Such requests may specify:

- 1) which items to examine
- 2) which tests to use
- 3) only source or sub-source level issues to be addressed.

The risk associated with these situations are, respectively,

- 1) examination of items that have not been submitted may provide stronger evidence in favour of either the prosecution or defence;

- 2) the potential benefits of other tests may be outside the knowledge of the customer;
- 3) opinion given at source or sub-source levels may be completely misleading when issues at higher levels (activity and offence) are being considered. The risk is double-sided: evidence that is evaluated as weak at source-level may in fact be stronger when considered at activity-level and, alternatively, evidence that is viewed as strong at source-level may in fact be much weaker at activity-level. In the first situation, the risk is that of guilty defendants being found not guilty because the scientific evidence is falsely weak; in the second situation, the risk is one of falsely convicting defendants when there is, in reality, insufficient evidence (e.g. as in the Barry George case).

Making decisions in-force on which items to submit or which tests to commission is probably viewed as the better option by the police as it gives them total control of submissions and budgets. Furthermore purchasing services at source and sub-source level is seen by customers as the more economic option. Activity-level services may involve more time and effort on the part of the provider and so are more expensive in the short term.

The CAI model encourages engagement of customers and providers in the Assessment phase in order to discuss, understand and agree an examination strategy that best meets the needs of the issues in the case. This is a sophisticated, business-like relationship that requires an element of trust on both sides of the transaction. It may be argued that, for whatever reasons, a low level of trust has developed since the

introduction of the commercial market in England and Wales and that this has led to the situation of prescriptive requests from customers.

In the long term, there is a price to pay for miscarriages of justice that is not currently included in any computations of value-for-money. Arguably, the situation can only get worse unless there is a strong counter-balance from effective defence challenge or from effective regulation. Neither of these counter-balances seems to be particularly strong, either currently or for the foreseeable future, with consequent implications for continuing miscarriages of justice.

Conclusions

The hypothesis of this thesis has been stated as:

The evaluation of a Bayesian likelihood ratio provides a philosophical and practical model to accommodate the potentially conflicting requirements of the criminal justice system for reliable, useful, expert opinion and of customers for economic, expert services.

A collection of papers has been presented in support of this hypothesis. The papers cover a period from 1998 to 2009 and provide an insight into the history of the development, firstly, of a basic, philosophical approach to assessment of a case and interpretation of evidence and, secondly, of the new insights gained from application of the model to various types of evidence and case.

Cook *et al.* (1998a) described for the first time a framework, called the Case Assessment and Interpretation (CAI) model, designed to guide scientists in making rational decisions in the face of pressures to use resources wisely and to deliver robust opinions.

The papers by Cook *et al.* (1998b), Evett, Jackson and Lambert (2000) and Cook *et al.* (1999) built on the basic model for Case Assessment and Interpretation (CAI). The novel notion of a hierarchy of propositions, the new distinction between explanations and propositions, and the logical interpretation of two-way transfer evidence that were described in these papers all gave fresh insight into the role and contribution of the forensic scientist.

The logical process of thought and of action that is captured by the CAI model can be translated into a logical structure for writing reports and statements of evidence. This structure was described in the paper by Evett *et al.* (2000a) and has influenced the way in which the main providers of forensic science in the United Kingdom, Republic of Ireland and parts of Europe write their reports and statements.

Application of the CAI approach to DNA-profiling and to drugs cases (Evett *et al.* 2002, Booth, Johnson and Jackson 2002) revealed new ways of appraising the value of these evidence types and, thereby, of presenting more reliable opinion to courts of law.

Clarity to the type of background data that are required to help experts interpret their findings robustly and justifiably was provided by the analysis of Champod, Evett and Jackson (2004). The paper provided for the first time a structured, logical framework within which to collect data rather than the somewhat *ad hoc* methods that had been formerly employed.

Jackson (2000), Jackson *et al.* (2006) and Jackson (2009) presented further development of the notion that there is a basic philosophical difference between the opinions offered by experts operating as 'investigators' or as 'evaluators'. This notion that there were two distinct roles for experts provided, for the first time, a clarity, both for the experts themselves and for their customers, on the contribution made by experts to the criminal justice process.

Finally, the paper by Jackson and Jones (2009) considered the application of CAI in the changing environment of commercial, and other, pressures and discussed some of the changes that would be necessary in order to maximise the benefits of CAI.

The papers that form the body of this thesis have been well received in the forensic science community, some having gained European recognition for their contribution to the development of forensic science, and are frequently quoted in the forensic science literature and in courts of law.

However, one recent judgment in the Court of Appeal (the case of *R v T*) has cast doubt on the admissibility of opinions based on likelihood ratios. Furthermore, commercial pressures on the police and other parties involved in the investigation of crime and in the prosecution and defence of people accused of crime run the risk of short-term approaches to decision-making that do not capitalise on the benefits of applying CAI.

On the other hand, the successful appeal of *Barry George* was largely due to a more robust appraisal of the weight of the evidence provided by a likelihood ratio approach. The standards published by the Association of Forensic Science Providers (AFSP 2009), a body comprising the main providers of forensic science in the UK and the Republic of Ireland, are based on the concepts described in the papers. The National Academy of Sciences in the United States of America has made recent, positive reference to some of the published papers.

It is this author's opinion that a 'likelihood ratio approach', because of its inherent logical structure, will not be displaced easily. Forensic science practitioners and lawyers who appreciate its benefits will keep CAI development alive, despite the trends or pressures not to do so, but only time will tell if the long-term benefits of CAI will be realised.

On balance, there is significant evidence to support the stated hypothesis.

References

- Aitken, C.G.G. 2009. Interpreting expert opinions: history of. In: A. Jamieson and A. Moenssens, eds. *Wiley Encyclopedia of Forensic Science*. Chichester, UK: John Wiley & Sons Ltd. 2009
- Aitken, C.G.G., Roberts, P., and Jackson, G. 2010. *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*. London: Royal Statistical Society
- Aitken, C.G.G. and Taroni, F. 1996 Interpretation of scientific evidence (Part II). *Science and Justice*. 37: 65
- Anderson, T., Schum, D. and Twining, W. 2005. *Analysis of Evidence*. Cambridge, UK: Cambridge University Press
- Association of Forensic Science Providers. 2009. Standards for the formulation of evaluative forensic science expert opinion. *Science and Justice*. 49: 161–164
- Bayes, T. 1763. An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*. 53: 370–418
- Berger, C., Buckleton, J., Champod, C., Evett, I.W. and Jackson, G., Evidence evaluation: a response to the Court of Appeal judgment in *R v T*. *Science and Justice*. [In press]

Booth, G, Johnston, F. and Jackson, G. 2002. Case assessment and interpretation in a drugs supply case. *Science and Justice*. 42: 123-125

Champod, C., Evett, I.W. and Jackson, G. 2004. Establishing the most appropriate databases for addressing source level propositions. *Science and Justice*. 44: 153-164

Cook R., Evett I.W., Jackson G., Jones P.J. and Lambert J.A. 1998a. A model for case assessment and interpretation. *Science and Justice*. 38: 151-156

Cook, R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. 1998b. A hierarchy of propositions: deciding which level to address in casework. *Science and Justice*. 38: 231-239

Cook R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. 1999. Case pre-assessment and review in a two-way transfer case. *Science and Justice*. 39: 103-111

Cooper, A. and Mason, L. 2009. Forensic resources and criminal investigations. In: J. Fraser and R. Williams, eds. *Handbook of Forensic Science*, Cullompton, Devon, UK: Willan Publishing. 2009, pp. 285-308

Craddock, J.G., Lamb, P. and Moffat, A.C. 1989. Problems of written communication: understanding and misunderstanding forensic scientists' statements. *Home Office Central Research and Services Establishment Technical Note 691*. London: Home Office

Curran, J.M., Hicks, T.N., and Buckleton, J.S. 2000. *Forensic Interpretation of Glass Evidence*. Boca Raton: CRC Press LLC

Davis, R., Facey, O., Hamer, P., and Rudram, D. 1996. Interpretation of scientific evidence. *Science and Justice*. 37: 64

ENFSI. 2000. [on-line]. Available from: <http://www.enfsi.eu/page.php?uid=29>
[Accessed 4 September 2010]

Evett, I.W. 1983. What is the probability that this blood came from that person? A meaningful question? *Journal of the Forensic Science Society*. 23: 35-39

Evett, I.W. 1984. A quantitative theory for interpreting transfer evidence in criminal cases. *Journal of the Royal Statistical Society (Series C - Applied Statistics)*. 33(1): 25-32

Evett, I.W. 1986. A Bayesian approach to the problem of interpreting glass evidence in forensic science casework. *Journal of the Forensic Science Society*. 26: 3-18

Evett, I.W. 1987a. Bayesian inference and forensic science: problems and perspectives. *The Statistician*. 36: 99-105

Evett, I.W. 1987b. On meaningful questions: the Bayesian approach to a two-trace problem. *Journal of the Forensic Science Society*. 27: 375-381

Evett, I.W. 1989. Some aspects of the Bayesian approach to evidence evaluation. *Journal of the Forensic Science Society*. 29: 303-316

Evett, I.W. 1990. The theory of interpreting scientific transfer evidence. *Forensic Science in Progress*. 4:141-179

Evett, I.W. 1991. Interpretation: a personal odyssey. In: C.G.G. Aitken and D.A. Stoney, eds. *The Use of Statistics in Forensic Science*. Chichester, West Sussex, UK: Ellis Horwood Limited. 1991, pp.9-22

Evett, I.W. 2011. Expressing evaluative opinions: a position statement. *Science and Justice*. 51: 1-2

Evett, I.W., Jackson, G. and Lambert, J.A. 2000. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Science and Justice*. 40: 3-10

Evett, I.W., Jackson, G., Lambert, J.A. and McCrossan, S. 2000a. The impact of the principles of evidence interpretation on the structure and content of statements. *Science and Justice*. 40: 233-239

Evett, I.W., Foreman, L.A., Jackson, G. and Lambert, J.A. 2000b. DNA profiling: a discussion of the issues relating to the reporting of very small match probabilities. *Criminal Law Review*. May: 341-355

Evetts, I.W., Gill, P.D., Jackson, G., Whitaker, J. and Champod, C. 2002. Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences*. 47: 520-530

Feinberg, S.E. 1997. Theories of legal evidence: what properties should they ideally possess and when are they informative? *The International Journal of Evidence and Proof*. 1 (SP1): 309-312

Finkelstein, M.O. and Fairley, W.B. 1970. A Bayesian approach to identification evidence. *Harvard Law Review*. 83: 489-517

Fraser, J. 2000. Not science...not support: forensic solutions to investigative problems. *Science and Justice*. 40: 127-130

Fraser, J. 2007. The application of forensic science to criminal investigation. In: T. Newburn, T. Williamson and A. Wright, eds. *Handbook of Criminal Investigation*. Cullompton, Devon, UK: Willan Publishing. 2007, pp. 381-402

Fraser, J. and Williams, R. 2009. The future of forensic investigations. In: J. Fraser and R. Williams, eds. *Handbook of Forensic Science*. Cullompton, Devon, UK: Willan Publishing. 2007, pp. 602-622

Gallop, A.M.C. 2003. Forensic science coming of Age. *Science and Justice*. 43: 55-59

Jackson, G. 2000. The scientist and the scales of justice. *Science and Justice*. 40: 81-85

Jackson, G. 2009. Understanding forensic science opinions. In: J. Fraser and R. Williams, eds. *Handbook of Forensic Science*. Cullompton, Devon, UK: Willan Publishing. 2009, pp. 419-455

Jackson, G, Jones, S., Booth, G., Champod, C. and Evett, I.W. 2006. The nature of forensic science opinion – a possible framework to guide thinking and practice in investigations and in court proceedings. *Science and Justice*. 46: 33-44

Jackson, G. and Jones, P.J. 2009. Case Assessment and Interpretation. In: A. Jamieson and A. Moenssens, eds. *Wiley Encyclopedia of Forensic Science*. Chichester, UK: John Wiley & Sons Ltd. 2009, pp. 483-497

Jeffreys, A.J., Wilson, V. and Thein, S.L. 1985. Hypervariable 'Minisatellite' Regions in Human DNA. *Nature*. 314: 67-73

Law Commission. 2011. *Expert Evidence in Criminal Proceedings in England and Wales* [on-line]. Available from: http://www.lawcom.gov.uk/expert_evidence.htm
[Accessed 23 March 2011]

Lawless, C. 2010. *A Curious Reconstruction? The Shaping of 'Marketized' Forensic Science*. London, UK: Centre for Analysis of Risk and Regulation, London School of Economics and Political Science

Lindley, D. 1977 A problem in forensic science. *Biometrika*. 64: 207-213

McCulloch, H. 1996. *Police use of forensic science*. Police Research Series Paper 19.
London: Home Office

Mullin, C. 1990. *Error of Judgement. The Truth about the Birmingham Bombings*.
Swords, Co Dublin, Ireland: Poolbeg Press Ltd.

National Academy of Sciences. 2009. *Strengthening Forensic Science in the United States: a Path Forward* [on-line].

Available from: http://www.nap.edu/openbook.php?record_id=12589&page=R1

[Accessed 21 February 2011]

National Policing Improvement Agency. 2010. '*National Forensic Framework Agreement*' [on-line]. Available from: <http://www.npia.police.uk/en/9398.htm>

[Accessed 1 March 2011]

Ramsay, M. 1987. *The Effectiveness of the Forensic Science Service*. Home Office
Research Study 92. London: Home Office

Redmayne, M., Roberts, P., Aitken, C.G.G. and Jackson, G. 2011. Forensic science
evidence in question. *Criminal Law Review*. Issue 5: 347-356

Rudram, D. A. 1996. Interpretation of scientific evidence. *Science and Justice*. 36: 133-138

Saks, M.J., Risinger, D.M., Rosenthal, R. and Thompson, W.C. 2003. Context effects in forensic science. *Science and Justice*. 43: 77-90

Satterthwaite, J. and Lambert, J. A. 1989. Interpreting the interpretations. A survey to assess the effectiveness of conclusions in statements written by forensic scientists. *Home Office Central Research and Services Establishment Technical Note 714*. London: Home Office

Taroni, F. and Aitken, C.G.G. 1996. Interpretation of scientific evidence. *Science and Justice*. 36: 290-292

Tilley, N. and Ford, A. 1996. *Forensic Science and Crime Investigation*. Crime Detection and Prevention Series Paper 73. London: Home Office

Tilley, N. and Townsley, 2009. Forensic Science in UK policing: strategies, tactics and effectiveness. In: J. Fraser and R. Williams, eds. *Handbook of Forensic Science*. Cullompton, Devon, UK: Willan Publishing. 2009, pp. 359-379

Touche Ross. 1987. *Review of Scientific Support for the Police*. London: Home Office

The 12 published papers cited below have been removed from the e-theses due to copyright restrictions.

Appendix 1: Published paper for Chapters 3

1.1 Cook R., Evett I.W., Jackson G., Jones P.J. and Lambert J.A. 1998a. A model for case assessment and interpretation. *Science and Justice*. 38: 151-156

Impact of this paper:

This paper has a total of 71 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=12634507729456642176&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

Appendix 2: Published papers for Chapter 4

2.1 Cook, R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. 1998b. A hierarchy of propositions: deciding which level to address in casework. *Science and Justice*. 38: 231-239

Impact of this paper:

This paper has a total of 98 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=40020918169816670&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

2.2 Evett, I.W., Jackson, G. and Lambert, J.A. 2000. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Science and Justice*. 40: 3-10

Impact of this paper:

This paper has a total of 48 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=12565172402821921912&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

2.3 Cook R., Evett, I.W., Jackson, G., Jones, P.J. and Lambert, J.A. 1999. Case pre-assessment and review in a two-way transfer case. *Science and Justice*. 39: 103-111

Impact of this paper:

This paper has a total of 36 citations from all sources on 12.05.11
(http://scholar.google.co.uk/scholar?cites=5913327983855151711&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

Appendix 3: Published paper for Chapter 5

3.1 Evett, I.W., Jackson, G., Lambert, J.A. and McCrossan, S. 2000a. The impact of the principles of evidence interpretation on the structure and content of statements. *Science and Justice*. 40: 233-239

Impact of this paper:

This paper has a total of 51 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=13549257956819422835&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

Appendix 4: Published papers for Chapter 6

4.1 Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J. and Champod, C. 2002. Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences*. 47: 520-530

Impact of this paper:

This paper has a total of 40 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=6217784199189695880&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

4.2 Booth, G, Johnston, F. and Jackson, G. 2002. Case assessment and interpretation in a drugs supply case. *Science and Justice*. 42: 123-125

Impact of this paper:

This paper has a total of 6 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=10842388061958902432&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

Appendix 5: Published paper for Chapter 7

5.1 Champod, C., Evett, I.W. and Jackson, G. 2004. Establishing the most appropriate databases for addressing source level propositions. *Science and Justice*. 44: 153-164

Impact of this paper:

This paper has a total of 7 citations from all sources on 12.05.11
(http://scholar.google.co.uk/scholar?cites=16971740023617199847&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

Appendix 6: Published papers for Chapter 8

6.1 Jackson, G. 2000. The scientist and the scales of justice. *Science and Justice*. 40: 81-85

Impact of this paper:

This paper has a total of 14 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=5984300855151728598&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

6.2 Jackson, G, Jones, S., Booth, G., Champod, C. and Evett, I.W. 2006. The nature of forensic science opinion – a possible framework to guide thinking and practice in investigations and in court proceedings. *Science and Justice*. 46: 33-44

Impact of this paper:

This paper has a total of 17 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=15267855381573319155&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

6.3 Jackson, G. 2009. Understanding forensic science opinions. In: J. Fraser and R. Williams, eds. *Handbook of Forensic Science*. Cullompton, Devon, UK: Willan Publishing. 2009, pp. 419-445

Impact of this paper:

This paper has a total of 2 citations from all sources on 12.05.11 (http://scholar.google.co.uk/scholar?cites=14957210255524981521&as_sdt=2005&sciodt=0,5&hl=en; accessed 12.05.11)

Appendix 7: Published paper for Chapter 9

7.1 Jackson, G. and Jones, P.J. 2009. Case Assessment and Interpretation. In: A. Jamieson and A. Moenssens, eds. *Wiley Encyclopedia of Forensic Science*. Chichester, UK: John Wiley & Sons Ltd. 2009, pp. 483-497

Impact of this paper:

This paper has 0 citations from all sources on 12.05.11 (<http://scholar.google.co.uk/>; accessed 12.05.11)